PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 17/00, C07K 7/04, 15/28, C12N 5/00, 15/00, C12P 21/06, C12Q 1/68

A1

(11) International Publication Number: WO 94/17087

(43) International Publication Date: 4 August 1994 (04.08.94)

(21) International Application Number: PCT/US94/01114

(22) International Filing Date: 28 January 1994 (28.01.94)

(30) Priority Data:

08/013,412	28 January 1993 (28.01.93)	US
08/087,119	30 June 1993 (30.06.93)	US

(81) Designated States: AU, CA, JP, KR, NZ, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, Oakland, CA 94612-3550 (US).

(72) Inventors: TJIAN, Robert; 964 Spruce Street, Berkeley, CA 94707 (US). COMAI, Lucio; 6707 Canyon Trail, El Cerrito, CA 94530 (US). DYNLACT, Brian, David; 42 8th Street, #1307, Charlestown, MA 02129 (US). HOEY, Timothy; 1423 Sanchez, San Francisco, CA 94131 (US). RUPPERT, Siegfried; 562 Spruce Street, Berkeley, CA 94707 (US). TANESE, Naoko; 464 Boynton Avenue, Berkeley, CA 94707 (US). WANG, Edith; 34655 Skylark Drive, #924, Union City, CA 94587 (US). WEINZIERL, Robert, O., J.; 1710 Walnut Street, #128, Berkeley, CA 94709 (US).

(74) Agents: ZIMMERMAN, C., Michael et al.; Flehr, Hohbach, Test, Albritton & Herbert, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).

(54) Title: TATA-BINDING PROTEIN ASSOCIATED FACTORS, NUCLEIC ACIDS ENCODING TAFS, AND METHODS OF USE

(57) Abstract

TATA-binding protein associated factors, TAFs, nuclear proteins involved in RNA polymerase I, II, and III transcription, and nucleic acids encoding TAFs are disclosed. The disclosed methods and compositions find use in developing pharmaceuticals, diagnosis and therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

TATA-BINDING PROTEIN ASSOCIATED FACTORS, NUCLEIC ACIDS ENCODING TAFS, AND METHODS OF USE

The research carried out in the subject application was supported in part by grants from the National Institutes of Health. The government may have rights in any patent issuing on this application.

CROSS-REFERENCE TO RELATED APPLICATION

This Application is a continuation-in-part of Application Serial No. 08/087,119 filed June 30, 1993, which is a continuation-in-part of Application Serial No. 08/013,412 filed January 28, 1993.

INTRODUCTION

Technical Field

The technical field of this invention concerns TATA-binding protein associated factors, proteins involved in gene transcription.

Background

Gene transcription requires the concerted action of a number of molecules. DNA provides regulatory sequences and a coding sequence, or template, from which an RNA polymerase synthesizes corresponding RNA. Regulatory sequences generally include sites for sequence-specific transcriptional control, including promoters, enhancers, suppressors, etc; and also a site for transcription initiation. For review, see Mitchell and Tjian (1989), Science 245, 371-378.

RNA polymerases alone appear incapable of initiating transcription. However, in vitro transcriptional activity of RNA polymerases can be restored by the addition of nuclear extracts or fractions thereof. For example, under certain conditions, in vitro transcription by RNA polymerase II (Pol II) can be at least partially restored by the addition of what have variously been reported to be four, five, six or seven nuclear fractions [See e.g. Matsui et al. (1980), Biol Chem 255, 1192], herein referred to as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIJ. Pol I and Pol III appear to require at least two fractions, called respectively SL1 and UBF, and TFIIIA and TFIIIB.

Many of these transcription fractions remain only partially characterized. For example, all but one of the Pol II fractions remain incompletely characterized or comprise multiple components. The fractions TFIID, SL1 and TFIIB have been reported to contain a TATA binding component, henceforth, TATA-binding protein, or TBP. Groups of the present Applicants have reported anti-TBP antibodies capable of immunoprecipitating TBP from TFIID, SL1, and TFIIIB.

TFIID, SL1 and TFIIIB immunoprecipitates have revealed TBP and numerous associated factors, tentatively called TBP-associated factors, or TAFs. Furthermore, preliminary experiments indicated that the TBP and non-TBP (TAF) fractions, when combined, facilitated at least some sequence-specific transcription activation.

Unfortunately, it is not clear from the above art that there is any transcriptional activity in the non-TBP fractions of TFIID, SL1 or TFIIB immunoprecipitates. For example, the reported apparent functional complementarity of the TBP and non-TBP fractions might result from the influence of antirepressors, inhibitor inhibition, etc. Furthermore, the coactivator transcriptional activity attributed to the non-TBP fractions could result from one or more components unrelated to the electrophoretically resolved TAF components. Nor does the literature provide any suggestion as to which, if any, of the electrophoretically resolved components of the non-TBP fraction provide(s) transcriptional activity, nor means for identifying bands resolvable from the non-TBP fractions.

Relevant Literature

Pugh and Tjian (1990), *Cell* 61:1187-1197; Tanese et al. (1991), *Genes and Devel* 5:2212-2224; Pugh and Tjian (1991), *Genes and Devel* 5:1935-1945; Dynlacht et al. (1991), *Cell* 66:563-576; Timmers et al. (1991), *Genes and Devel* 5:1946-1956; Zhou et al. (1992), *Genes and Devel* 6:1964-1974; and Takada et al. (1992), *Proc Natl Acad Sci USA* 89:11809-11813, relate to factors associated with Pol II transcription. Comai et al. (1992) *Cell* 68:965-976 relates to factors associated with Pol I transcription. Lobo et al. (1991), *Genes and Devel*, 5:1477-1489; Margotin et al. (1991), *Science* 251:424-426; Simmen et al. (1991), *EMBO J* 10:1853-1862; and Taggart et al. (1992), *Cell* 71:1015; Lobo et al. (1992), *Cell* 71:1029; and White and Jackson (1992), *Cell* 71:1041 relate to factors associated with Pol III transcription. Sekiguchi et al. (1988), *EMBO J* 7:1683-1687 and Sekiguchi et al. (1991), *Mol and Cellular Biol* 11:3317-3325 disclose the cloning of the CCG1 gene encoding a protein reported to be involved in cell cycle progression.

SUMMARY OF THE INVENTION

Substantially pure and biologically active TATA-binding protein associated factors (TAFs), eukaryotic nuclear proteins involved in RNA polymerase I, II, and III transcription, nucleic acids encoding TAFs, and methods of using TAFs and TAF-encoding nucleic acids are provided. Recombinant TAFs, anti-TAF antibodies and TAF-fusion products find use in drug screening, diagnositcs and therapeutics. In particular, the disclosed TAFs provide valuable reagents in developing specific biochemical assays for screening compounds that agonize or antagonize selected transcription factors involved in regulating gene expression associated with human pathology.

DESCRIPTION OF SPECIFIC EMBODIMENTS

Substantially pure and biologically active TATA-binding protein associated factors (TAFs) and portions thereof, nucleic acids encoding TAFs and portions thereof, and methods of use are provided.

As used herein, a given TAF refers to the TAF protein, recombinant or purified from a natural source, and functional and xenogeneic analogs thereof. For

example "dTAFII110" refers to a Pol II TAF, deriveable from Drosophila, with an apparent molecular weight of about 110 kD, generally as determined by SDS-PAGE under conditions described herein, in Dynlacht et al. (1991), Comai et al. (1992), or otherwise identified by functional, sequence, etc. data herein. It is understood that these molecular weight designations are for the convenience of nomenclature and may not necessarily correspond to actual or predicted molecular weight. Other TAFs are analogously identified herein.

A "portion" of a given TAF is a peptide comprising at least about a six, preferably at least about an eighteen, more preferably at least about a thirty-six amino acid sequence of the TAF. Of particular interest are portions of the TAF that facilitate functional or structural interaction with activators, TAFs, TBP, Pol I, II or III, the TATA box and surrounding DNA sequences, etc. Methods for identifying such preferred portions are described below.

By substantially full-length is meant a polypeptide or polynucleotide that comprises at least 50%, preferably at least 70% and more preferably at least 90% of the natural TAF polypeptide or polynucleotide length.

"Xenogeneic" TAF analogs are nonhuman-, nonDrosophila-derived proteins with substantial functional or sequence identity to human and Drosophila TAFs. Of particular interest are xenogeneic TAF analogs derived from rodents, primates, and livestock animals including bovine, ovine, equine and avian species

"Functional" analogs of a given TAF or proteins with "substantial functional identity" to a given TAF are compounds that exhibit one or more biochemical properties specific to such TAF, such as the ability of dTAFII110 to interact with Sp1.

"Modulating transcription" means altering transcription, and includes changing the rate of transcription initiation, the level of transcription, or the responsiveness of transcription/transcription initiation to regulatory controls.

The terms "substantially pure" or "isolated" mean that the TAF, TAF portion, or nucleic acid encoding a TAF or TAF portion is unaccompanied by at least some of the material with which it is normally associated in its natural state. While a composition of a substantially pure TAF or portion thereof is preferably substantially free of polyacrylamide, such composition may contain excipients and additives useful in diagnostic, therapeutic and investigative reagents. A

substantially pure TAF composition subject to electrophoresis or reverse phase HPLC provides such TAF as a single discernable proteinaceous band or peak.

Generally, a substantially pure TAF composition is at least about 1% protein weight said TAF; preferably at least about 10%; more preferably at least about 50%; and most preferably at least 90%. Protein weight percentages are determined by dividing the weight of the TAF or TAF portion, including alternative forms and analogs of the TAF such as proteolytic breakdown products, alternatively spliced, differentially phosphorylated or glycosylated, or otherwise post-translationally modified forms of the TAF, present in a fraction by the total protein weight present.

A biologically active TAF or TAF portion retains one or more of the TAF's native function such as the ability to specifically bind TBP, transcription factors (activators), other TAFs or anti-TAF antibodies, or to modulate or facilitate transcription or transcription initiation. Exemplary assays for biological activity are described below and in the working exemplification.

Specific binding is empirically determined by contacting, for example a TAF, with a mixture of components and identifying those components that preferentially bind the TAF. Specific binding may be conveniently shown by competitive binding studies, for example, immobilizing a TAF, on a solid matrix such as a polymer bead or microtiter plate and contacting the immobilized TAF with a mixture. Often, one or more components of the mixture will be labelled. Another useful approach is to displace labelled ligand. Generally, specific binding of a TAF will have binding affinity of 10^{-6}M, preferably 10^{-8}M, more preferably 10^{-10}M under optimized reaction conditions and temperature.

Portions of TAFs find use in screening TAF expression libraries, defining functional domains of TAFs, identifying compounds that bind or associate with TAFs and the like. Accordingly, peptides encoding a portion of a TAF are provided that are capable of modulating transcription including transcription initiation. Typically, such peptides are effective by binding to a TAF, an activator, or TBP or competitively inhibiting a TAF domain's association with another compound, typically a protein like TBP or another TAF, an activator, or DNA. For example, TAF-TAF interactions may be exploited to purify TAFs, e.g. immobilized TAF200 is used to purify TAF110.

Associational domains of TAFs are ascertainable by those skilled in the art using the methods and compositions disclosed herein. Useful methods include in vitro mutagenesis such as deletion mutants, secondary and tertiary structural predictions, antibody and solvent accessibility, etc. For example, peptides derived from highly charged regions find particular use as immunogens and as modulators of TAF-protein interactions. Also, TAF mutants are used to identify regions important for specific protein interactions or otherwise involved in transcription. Here, useful assays include column binding assay and transfection studies.

The invention provides recombinantly produced TAFs, TAF analogs and portions thereof. These recombinant products are readily modified through physical, chemical, and molecular techniques disclosed or cited herein or otherwise known to those skilled in the relevant art. According to a particular embodiment of the invention, portions of the TAF-encoding sequences are spliced with heterologous sequences to produce fusion proteins. Such fusion proteins find particular use in modulating gene transcription in vitro and in vivo.

For example, many eukaryotic sequence-specific transcription factors have separable DNA binding and activation domains. A TAF or domain thereof can be fused to a well-characterized DNA binding domain (see, e.g., Sadowski et al., (1988) Nature 335, 563-564) and the resulting fusion protein can be tested for its ability to modulate transcription or transcriptional initiation. For example, we disclose the fusion of the N-terminal region of TAF110 to the DNA binding domain of the GAL4 protein. Alternatively, an TAF domain can be fused with a domain having endonuclease activity for site-specific DNA cleaving. Other useful TAF fusion partners include GST, Lerner epitope, an epitope recognized by a monoclonal antibody (e.g. hemagglutinin epitope and 12CA5 monoclonal antibody), glutathione S-transferase for immobilization, the SP1 or VP16 activation domains, etc.

TAFs can be further modified by methods known in the art. For example, TAFs may be phosphorylated or dephosphorylated, glycosylated or deglycosylated, with or without radioactive labeling, etc. The disclosed TAF serine residues in particular provide useful phosphorylation sites. See e.g. methods disclosed in Roberts et al. (1991) Science 253, 1022-1026 and in Wegner et al. (1992) Science 256, 370-373. Especially useful are modifications that alter TAF solubility,

membrane transportability, stability, and binding specificity and affinity. Some examples include fatty acid-acylation, proteolysis, and mutations in TAF-TAF or TAF-TBP interaction domains that stabilize binding.

TAFs may also be modified with a label capable of providing a detectable signal, for example, at a heart muscle kinase labeling site, either directly or indirectly. Exemplary labels include radioisotopes, fluorescers, etc. Alternatively, a TAF may be expressed in the presence of a labelled amino acid such as ^{35}S-methionine. Such labeled TAFs and analogs thereof find use, for example, as probes in expression screening assays for proteins that interact with TAFs, or, for example, TAF binding to other transcription factors in drug screening assays.

Specific polyclonal or monoclonal antibodies that can distinguish TAFs from other nuclear proteins are conveniently made using the methods and compositions disclosed in Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, 1988, other references cited herein, as well as immunological and hybridoma technologies known to those in the art. In particular, TAFs and analogs and portions thereof also find use in raising anti-TAF antibodies in laboratory animals such as mice and rabbits as well as the production of monoclonal antibodies by cell fusion or transformation.

Anti-TAF antibodies and fragments (Fab, etc) thereof find use in modulating TAF involvement in transcription complexes, screening TAF expression libraries, etc. In addition, these antibodies can be used to identify, isolate, and purify structural analogs of TAFs. Anti-TAF antibodies also find use for subcellular localization of TAFs under various conditions such as infection, during various cell cycle phases, induction with cytokines, protein kinases such as C and A, etc. Other exemplary applications include using TAF-specific antibodies (including monoclonal or TAF-derived peptide specific antibodies) to immuno-deplete in vitro transcription extracts and using immuno-affinity chromatography to purify TAFs, including analogs, or other nuclear factors which interact with TAFs.

Compositions are also provided for therapeutic intervention in disease, for example, by modifying TAFs or TAF encoding nucleic acids. Oligopeptides can be synthesized in pure form and can find many uses in diagnosis and therapy. These oligopeptides can be used, for example, to modulate native TAF interaction with other TAFs, TBP, other transcription factors or DNA. The oligopeptides will

generally be more than six and fewer than about 60 amino acids, more usually fewer than about 30 amino acids, although large oligopeptides may be employed. A TAF or a portion thereof may be used in purified form, generally greater than about 50%, usually greater than about 90% pure. Methods for purifying such peptides to such purities include various forms of chromatographic, chemical, and electrophoretic separations disclosed herein or otherwise known to those skilled in the art.

Experimental methods for purifying TAFs are set out briefly below and in detail in the following working exemplification. Generally, TBP-TAF complexes are immunopurified (generally, by immunoprecipitation) using polyclonal or monoclonal antibodies directed against a native TAF or TBP epitope. Alternatively, monoclonal antibodies directed against an epitope-tagged TBP or TAF may be used. See e.g. Zhou, et al. (1992). At least three complementary experimental approaches are employed for isolating cDNAs encoding TAFs: (1) TAF-specific binding proteins (e.g. antibodies directed against TAF proteins, TAF-binding TAFs, TBP, TAF-binding activators, or TAF-binding coactivators) are used for screening expression libraries; (2) cDNA libraries are screened with potentially homologous TAF oligonucleotide sequences (alternatively, a series of degenerate oligonucleotide PCR primers derived from the homologous TAF sequence may be used to amplify probes from cDNA. See Peterson et al. (1990) Science, 248, 1625-1630, Figure1.); and, (3) TAF proteins are purified to homogeneity for protein microsequencing.

TAF ENCODING NUCLEIC ACID

The invention provides nucleic acid sequences encoding TAFs and portions of TAFs. By "encoding a portion of a TAF" is meant to include sequences substantially identical to sequences encoding at least a portion of a TAF. Included are DNA and RNA sequences, sense and antisense.

"Substantial sequence identity" means that a portion of the protein or nucleic acid presents at least about 70%, more preferably at least about 80%, and most preferably at least about 90% sequence identity with a TAF sequence portion. Where the sequence diverges from native TAF sequences disclosed herein, the differences are preferably conservative, i.e. an acidic for an acidic amino acid

substitution or a nucleotide change providing a redundant codon. Dissimilar sequences are typically aggregated within regions rather than being distributed evenly over the polymer.

A substantially identical sequence hybridizes to a complementary TAF-encoding sequence under low stringency conditions, for example, at 50°C and 6X SSC (0.9M saline/0.09M sodium citrate) and that remains bound when subject to washing at 55°C with 1X SSC.

The invention's TAF encoding polynucleotides are isolated; meaning that the claimed sequence is present as other than a naturally occurring chromosome or transcript in its natural environment. Typically isolated sequences are removed from at least some of the nucleotide sequences with which they are normally associated with on a natural chromosome.

A substantially pure or isolated TAF- or TAF portion-encoding nucleic acid is generally at least about 1% nucleic acid weight said TAF-encoding nucleic acid; preferably at least about 10%; more preferably at least about 50%; and most preferably at least 90%. Nucleic acid weight percentages are determined by dividing the weight of the TAF or TAF portion-encoding nucleic acid, including alternative forms and analogs such as alternatively spliced or partially transcribed forms, by the total nucleic acid weight present.

The invention also provides for TAF sequences modified by transitions, transversions, deletions, insertions, or other modifications such as alternative splicing and such alternative forms, genomic TAF sequences, TAF gene flanking sequences, including TAF regulatory sequences and other non-transcribed TAF sequences, TAF mRNA sequences, and RNA and DNA antisense sequences complementary to TAF encoding sequences, equences encoding xenogeneic TAFs, and TAF sequences comprising synthetic nucleotides, e.g., the oxygen of the phosphate group may be replaced with sulfur, methyl, or the like.

For modified TAF-encoding sequences or related sequences encoding proteins with TAF-like functions, there will generally be substantial sequence identity between at least a portion thereof and a portion of a TAF, preferably at least about 40%, more preferably at least 80%, most preferably at least 90%, particularly conservative substitutions, particularly within regulatory regions and

regions encoding protein domains involved in protein-protein interactions, particularly TAF-transcription factor interactions.

Typically, the invention's TAF encoding polynucleotides are associated with heterologous sequences. Examples of such heterologous sequences include regulatory sequences such as promoters, enhancers, response elements, signal sequences, polyadenylation sequences, etc., introns, 5' and 3' noncoding regions, etc. Other useful heterologous sequences are known to those skilled in the art or otherwise disclosed references cited herein. See for example, Russel Doolittle, Of URFs and ORFs, A Primer on How to Analyze Derived Amino Acid Sequences, University Science Books, Mill Valley CA.

TAF encoding nucleic acids can be subject to alternative purification, synthesis, modification or use by methods disclosed herein or otherwise known in the art. For example, the nucleic acids can be modified to alter stability, solubility, binding affinity and specificity, methylation, etc. The nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescers, biotinylation, etc.

Nucleic acids encoding at least a portion of a TAF are used to identify nuclear factors which interact with that TAF using expression screening in yeast as described in Current Protocols in Molecular Biology. In this example, a yeast cDNA library containing fusion genes of cDNA joined with DNA encoding the activation domain of a transcription factor (e.g. Gal4) are transfected with fusion genes encoding a portion of a TAF and the DNA binding domain of a transcription factor. Clones encoding TAF binding proteins provide for the complementation of the transcription factor and are identified through transcription of a reporter gene. See, e.g. Fields and Song (1989) Nature 340, 245-246 and Chien et al. (1991) Proc Natl Acad Sci USA 88, 9578-9582.

The invention also provides vectors comprising nucleic acids encoding a TAF or portion or analog thereof. A large number of vectors, including plasmid and viral vectors, have been described for expression in a variety of eukaryotic and prokaryotic hosts. Vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g. antibiotic resistance, and one or more expression cassettes. The inserted TAF coding

sequences may be synthesized, isolated from natural sources, prepared as hybrids, etc. Ligation of the coding sequences to the transcriptional regulatory sequences may be achieved by known methods. Advantageously, vectors may also include a promotor operably linked to the TAF encoding portion.

Suitable host cells may be transformed/transfected/infected by any suitable method including electroporation, $CaCl_2$ mediated DNA uptake, viral infection, microinjection, microprojectile, or other established methods. Alternatively, nucleic acids encoding one or more TAFs may be introduced into cells by recombination events. For example, a sequence can be microinjected into a cell, and thereby effect homologous recombination at the site of an endogenous gene encoding a TAF, an analog or pseudogene thereof, or a sequence with substantial identity to a TAF-encoding gene. Other recombination-based methods such as nonhomologous recombinations, deletion of endogenous gene by homologous recombination, especially in pluripotent cells, etc., provide additional applications.

Appropriate host cells include bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are E. coli, B. subtilis, Saccharomyces cerevisiae, SF9 and SF21 cells, C129 cells, 293 cells, Neurospora, and CHO, COS, HeLa cells and immortalized mammalian myeloid and lymphoid cell lines. Preferred replication systems include M13, ColE1, SV40, baculovirus, vaccinia, lambda, adenovirus, AAV, BPV, etc. A large number of transcription initiation and termination regulatory elements/regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art. The particular choice of vector/host cell is not critical to the invention.

Under appropriate expression conditions, host cells are used as a source of recombinantly produced TAFs or TAF analogs. Preferred expression systems include E. Coli, vaccinia, or baculovirus; the latter two permitting the recombinant TAFs to be modified, processed and transported within a eukaryotic system.

TAF-encoding oligonucleotides also used to identify other TAFs or transcription factors. For example, ^{32}P-labeled TAF-encoding nucleic acids are used to screen cDNA libraries at low stringency to identify similar cDNAs that encode proteins with TAF-related domains. Additionally, TAF related proteins are

isolated by PCR amplification with degenerate oligonucleotide probes using the sequences disclosed herein. Other experimental methods for cloning TAFs, sequencing DNA encoding TAFs, and expressing recombinant TAFs are also set out in the working exemplification below. Other useful cloning, expression, and genetic manipulation techniques for practicing the inventions disclosed herein are known to those skilled in the art.

The compositions and methods disclosed herein may be used to effect gene therapy. See, e.g. Gutierrez et al. (1992) Lancet 339, 715-721. For example, cells are transfected with TAF sequences operably linked to gene regulatory sequences capable of effecting altered TAF expression or regulation. To modulate TAF translation, cells may be transfected with TAF complementary antisense polynucleotides.

Antisense modulation may employ TAF antisense sequences operably linked to gene regulatory sequences. Cells are transfected with a vector comprising a TAF sequence with a promoter sequence oriented such that transcription of the gene yields an antisense transcript capable of binding to TAF encoding mRNA. Transcription may be constitutive or inducible and the vector may provide for stable extrachromosomal maintenance or integration. Alternatively, single-stranded antisense nucleic acid sequences that bind to genomic DNA or mRNA encoding at least a portion of TAF may be administered to the target cell at a concentration that results in a substantial reduction in TAF expression.

ASSAYS FOR IDENTIFYING TRANSCRIPTION FACTORS AND THERAPEUTIC AGENTS

The invention provides methods and compositions for identifying agents useful in modulating gene transcription. Such agents find use in the diagnosis or treatment of broad range of disease including cancer, cardiovascular diseases, microbial and fungal infections and particularly viral infections, inflammatory disease, immune disease, etc. The ability to develop rapid and convenient high-throughput biochemical assays for screening compounds that interfere with the process of transcription in human cells opens a new avenue for drug development. An overview of this therapeutic approach is presented in Peterson & Baichwal (1993), Trends in Biotechnology, in press.

Typically, prospective agents are screened from large libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of saccharide, peptide, and nucleic acid based compounds, see, e.g. Lam et al., (1991) Nature 354, 82-86. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily predicable. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means. Examples of such modifications are disclosed herein.

Useful agents are identified with a range of assays employing TAFs or TAF encoding nucleic acids. As examples, protein binding assays, nucleic acid binding assays and gel shift assays are useful approaches. Exemplary assays include assaying labeled TBP binding to immobilized TAF, labeled TAF or TAF peptide binding immobilized TBP, etc. Many appropriate assays are amenable to scaled-up, high throughput usage suitable for volume drug screening. Such screening will typically require the screening of at least about 10, preferably at least about 100, and more preferably at least about 1000 prospective agents per week. The particular assay used will be determined by the particular nature of the TAF interactions. For instance, a prospective agent may modify with the function of a TAF but not with transcription complex assembly. For example, a molecule that binds to a TAF but does not disrupt complex assembly is identified more readily through labelled binding assays than through gel retardation assay. Assays may employ single TAFS, TAF portions, TAF fusion products, partial TAF complexes, or the complete TFIID transcription complex, depending on the associational requirements of the subject transcription factor.

Useful agents are typically those that bind to or modify the association of transcription associated factors, especially TAFs. Preferred agents include those capable of modulating the expression of Pol II genes, particulary oncogenes (including viral oncogenes such as adenovirus EIA, human papilloma E7, and cellular oncogenes such as Rb, P53, E2F, myc, fos/jun (AP1), abl, etc.), genes transcribed during viral infection or activation, and sterol regulated genes. Preferered agents modify, preferably disrupt, TAF-TAF, TAF-activator, TAF-coactivator (coactivators include OCA-B, dTAFII110, etc.) or TAF-TBP binding. An especially preferred useful agent disrupts the association of a disclosed hTAF,

with an activator, particularly a viral-specific activator, particularly an HIV-specific activator such as tat.

Useful agents are found within numerous chemical classes, though typically they are organic compounds; preferably small organic compounds. Small organic compounds have a molecular weight of more than 50 yet less than about 2,500, preferably less than about 750, more preferably, less than about 250. Exemplary classes include peptides, saccharides, steroids, and the like.

Selected agents may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways to enhance their stability, such as using an unnatural amino acid, such as a D-amino acid, particularly D-alanine, by functionalizing the amino or carboxyl terminus, e.g., for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like. Other methods of stabilization may include encapsulation, for example, in liposomes, etc.

Agents may be prepared in a variety of ways known to those skilled in the art. For example, peptides under about 60 amino acids can be readily synthesized today using conventional commercially available automatic synthesizers. Alternatively, peptide (and protein and nucleic acid agents) are readily produced by known recombinant technologies.

For therapeutic uses, the compositions and selected agents disclosed herein may be administered by any convenient way that will depend upon the nature of the compound. For small molecular weight agents, oral administration is preferred and enteric coatings may be indicated where the compound is not expected to retain activity after exposure to the stomach environment. Generally the amount administered will be empirically determined, typically in the range of about 1 to 1000 ug/kg of recipient.

Large proteins are preferably administered parenterally, conveniently in a physiologically acceptable carrier, e.g., phosphate buffered saline, saline, deionized water, or the like. Typically, such compositions are added to a retained physiological fluid such as blood or synovial fluid. Generally, the amount administered will be empirically determined, typically in the range of about 10 to

1000 μg/kg of the recipient. Other additives may be included, such as stabilizers, bactericides, etc. These additives will be present in conventional amounts.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Additional exemplary materials and methods for the purification, cloning and expression of TAFs are described below. Additional exemplary functional assays are described in detail. While exemplified primarily for dTAFII110, the disclosed methods find ready application to other TAFs by those skilled in the art and familiar with the methods hereinor found in standard manuals such as Molecular Cloning, A Laboratory Manual (2nd Ed., Sambrook, Fritsch and Maniatis, Cold Spring Harbor), Current Protocols in Molecular Biology (Eds. Ausubel, Brent, Kingston, Moore, Seidman, Smith and Struhl, Greene Publ. Assoc., Wiley-Interscience, NY, NY, 1992)

Immunopurified dTFIID complex is necessary and sufficient to mediate Sp1 activation in vitro.

In order to determine if the TFIID complex is sufficient to substitute for a partially-purified TFIID fraction, we have purified the TBP-TAF complex extensively by using an affinity resin coupled to a specific monoclonal antibody directed against TBP. Transcriptionally active TFIID purified from Drosophila embryos was obtained by eluting the complex from the antibody affinity resin with a low concentration (0.5 M) of guanidine hydrochloride in the presence of a synthetic peptide corresponding to the epitope recognized by monoclonal 42A11. The antibody used for the immunopurification remained bound to the protein G-sepharose beads and was found in the pellet. The proteins were electrophoresed on an 8 % polyacrylamide-SDS gel and detected by silver staining. The resultant gels reveal seven major TAFs in the complex ranging in size from 30 to over 200 kD.

After dialysis of the purified TFIID complex to remove the peptide and denaturant, in vitro transcription reactions were carried out in the presence of basal factors that were isolated from Drosophila embryo nuclear extracts (Dynlacht et al., 1991; Wampler et al., 1990). Without the TFIID fraction there is no

detectable transcription. Purified, recombinant dTBP is able to direct basal but not activated transcription. In contrast, immunopurified TFIID complex is able to mediate basal expression and Sp1 activation. Sp1-dependent activation with the TFIID fraction is shown in lanes 7 and 8. For the in vitro transcription assay, 2 ul of the immunopurified TFIID complex was assayed. Transcription was assayed by primer extension. The results demonstrate that the immunopurified TFIID complex containing TBP and at least 7 specific TAFs is necessary and sufficient for Sp1-dependent activation of transcription in vitro. As expected, the impure TFIID fraction also mediates transcriptional activation by Sp1, while the recombinant TBP protein is only able to direct basal, but not activated transcription. The immunopurified complex is also able to support activation by other transcription factors such as NTF-1.

Cloning and expression of Drosophila TAF110 cDNAs

Purified TFIID complex was used to immunize a mouse, and monoclonal antibodies were generated against TAF110 (see Experimental Procedures below). The serum from the immunized mouse was also collected and polyclonal antibodies used to screen a λgt11 expression library constructed from Drosophila embryo cDNA (Zinn et al., 1988). One clone was tentatively classified as a TAF110 cDNA because it produced protein that cross-reacted with independently isolated anti-TAF110 monoclonal antibodies. This partial cDNA clone was subsequently used as a probe to isolate full-length cDNAs from a λgt10 library (Poole et al.,1985). The longest clone obtained was 4.6 kb. This cDNA is polyadenylated at the 3' end and appears to be nearly full-length, based on the size of the mRNA, as determined by Northern blot analysis. The 4.6 kb cDNA clone contains a long open reading frame coding for a protein of 921 amino acids (SEQUENCE ID NO:1), with a calculated molecular weight of 99.4 kD and an estimated pI of 10.1. Within the predicted amino acid sequence, there are 3 peptides That correspond to amino acid sequences determined from lys C peptides generated from HPLC purified TAF110. For microsequencing, the TFIID complex was immunopurified from fractionated embryo nuclear extract, and the TAFs were separated from TBP and the antibody by elution with 1 M guanidine-HCl. The purified TAFs were fractionated on a C4 reverse phase HPLC column. Three adjacent fractions

containing TAF 110 as the major species were cleaved with the protease lys-C, and the resulting peptides were purified and sequenced. Three peptide sequences were found that match the predicted amino acid sequence of the TAF110 cDNA

We have expressed TAF 110 protein in a variety of cell types. The protein was expressed from the cloned gene in a baculovirus expression system and detected by western blot using a TAF110 monoclonal antibody. The protein encoded by the TAF110 cDNA has the same apparent molecular weight as the endogenous protein in the TFIID fraction derived from Drosophila cells, and the protein produced from the cloned gene cross-reacts with monoclonal antibodies directed against the TAF110 protein isolated from embryos. These results taken together demonstrate that the 4.6 kb cDNA encodes the full-length TAF110 protein.

TAF110 appears to be a single copy gene in Drosophila based on low-stringency Southern blot analysis. The TAF110 gene is located at 72D,4-5 on the left arm of the third chromosome. There are not any previously identified Drosophila genes assigned to this chromosomal location (Lindsley and Zimm, 1992).

Hybridomas producing antibodies against TAF110 were selected by screening cell culture supernatants for those containing antibodies that specifically recognize the 110 kd protein in a western blot. For westerns, approximately 50 ug of the TFIID fraction was immunoprecipitated with antibodies against dTBP or TAF110. The α-TAF110 monoclonal antibody 33G8 was obtained from a hybridoma culture medium and purified by binding to protein G-sepharose. Proteins were eluted from the resin by boiling in sample buffer, electrophoresed on 8% polyacrylamide gel, and silver stained. Several of the a-TAF110 monoclonals that were obtained by this method specifically immunoprecipitate the same set of proteins as a-dTBP antibodies. This demonstrates that at least part of TAF110 is accessible to our antibodies, and therefore exposed in the native TFIID complex and positioned for interaction with activators.

Monoclonal antibodies specific for other Drosophila TAFs can also immunopurify the same TFIID complex as a-TBP and α-TAF110 antibodies. Thus, there appears to be one predominant TBP-containing complex in the TFIID fraction, as opposed to a heterogeneous set of complexes containing different sets

of TAFs bound to TBP. Our methods are also used to determine if there are rare, perhaps tissue-specific, TBP-containing complexes that might contain different collections of TAFs or if the activity of the TAFs could be modulated by post-translational modifications. For example, TAF200 does not stain as intensely as the other TAFs and TBP, and, on this basis, might not be present in all complexes. However, this protein seems to be an authentic member of the major TFIID complex since antibodies directed against TAF200 immunopurify a set of proteins that appear to be identical to complexes purified by antibodies directed against TBP or other TAFs. The preparations of the purified TFIID complex contain some polypeptides that are less abundant than the major TAF proteins. Based on western analyses with α-TAF antibodies, these minor species appear to be proteolytic breakdown products of larger TAFs or substoichiometric TAFs.

The TAF110 coding sequence contains several regions which are rich in glutamine residues or rich in serine and threonine residues, and the C-terminal third of the protein is highly charged. The C-terminal region of the molecule contains 32% acidic or basic residues. We searched the existing data bases for genes similar to the TAF110 gene, and found that it is not highly homologous to any previously identified genes. In particular, TAF110 did not show any similarity to any DNA binding domains. Interestingly, Sp1 received one of the highest scores in the NBRF protein sequence data base search for similarity to TAF110. The amino terminal third of TAF110 has an organization similar to the activation domains of Sp1, consisting of glutamine-rich regions flanked by serine-threonine rich domains. The two proteins share 21% amino acid identity and 35% similarity over 260 residues.

This unexpected similarity to Sp1 prompted us to consider a possible functional relationship between Sp1 and TAF110. In particular, whether the amino-terminal region of TAF110 might contain interaction surfaces for activators such as Sp1, especially since the A and B glutamine-rich domains are responsible for mediating Sp1-Sp1 interactions as well as activation. Indeed, one of the unique properties of Sp1 activation domains is their capacity to mediate a phenomenon called superactivation, in which a truncated form of Sp1 lacking the zinc fingers but containing glutamine-rich domains A and B is able to interact directly with DNA-bound full length Sp1. This interaction increases the number of activation

domains at the promoter and can greatly enhance expression of a gene regulated by Sp1 binding sites. This type of interaction also appears to be involved in synergistic activation mediated by distally and proximally bound Sp1.

?dTAF110 can function as a target for the Sp1 activation domains

To test for functional homology between the similar domains, we asked if the N-terminal region of TAF110 could function as a target for the Sp1 activation domains in a superactivation assay. The amino terminal 308 residues of TAF110 were fused to the DNA binding domain of the GAL4 protein, G4(1-147), and tested in a transient cotransfection assay in Drosophila Schneider cells. This hybrid construct, by itself, weakly activates (4 fold) a reporter gene which is dependent on GAL4 binding sites. This low level of activity is similar to the modest activation observed with constructs containing the Sp1 B domain fused to GAL4. When this TAF110 hybrid construct is cotransfected with DNA expressing the gln-rich A and B domains of Sp1, (N539), a 60 fold increase in transcription is observed. This 15 fold superactivation is dependent on the TAF110 sequences since Sp1(N539) is unable to stimulate transcription when cotransfected with G4(1-147) alone. The interaction with Sp1 apparently requires an extended region of TAF110 since GAL4 fusion proteins bearing TAF110 residues 1-137, 138-308, or 87-308 are unable to mediate superactivation by Sp1.

These results indicate that the N-terminal 308 amino acids of TAF110 are sufficient for mediating an interaction with the glutamine-rich activation domains of Sp1 that lead to superactivation. In the positive control for this experiment, a GAL4-Sp1B domain fusion is superactivated approximately 50 fold by the fingerless Sp1 mutant. In a search for other potential targets of Sp1, we have tested some additional members of the TFIID complex for the ability to mediate superactivation by Sp1. For example, GAL4 hybrids containing TAF40, TAF80, or the amino-terminal region of dTBP were found to be inactive in the superactivation assay. This results shows that the interaction between TAF110 and Sp1 in Drosophila cells is quite specific and that other subunits of the TBP-TAF complex that we tested are unable to interact with the glutamine-rich activation domains of Sp1.

dTAF110 and Sp1 interact in yeast

The superactivation assay in Drosophila Schneider cells provided the first hint that TAF110 may serve as a coactivator for Sp1. However, it is difficult to assess in this assay whether TAF110 can interact with Sp1 in the absence of the other TAFs which are present in Drosophila cells. The superactivation assay also imposes certain limitations to the number and types of constructs that can be tested. Moreover, it seemed prudent to establish several independent assays to investigate the relationship between TAF110 and transcription activation domains. Therefore, we carried out two additional types of assays, one in vivo and one in vitro, to test the results obtained in Schneider cells. First, we tested the ability of TAF110 and Sp1 to interact in a versatile assay for protein-protein interaction which is carried out in yeast cells (Fields and Song, 1989). This strategy takes advantage of the modular organization of eukaryotic transcription factors. In this assay, one of the partners to be tested is fused to the DNA binding domain of GAL4 and, in a separate molecule, the other partner is fused to the acidic activation domain (AAD). A functional activation domain is recruited to the target promoter bearing GAL4 binding sites and the lacZ reporter gene is expressed only if there is a protein-protein interaction between the partners being tested.

Full-length TAF110 as well as a variety of deletion mutants were fused to the DNA binding domain of GAL4, G4(1-147). In contrast to the situation in Drosophila cells, the amino terminal region of TAF110 cannot activate transcription by itself in yeast. This result was anticipated since glutamine-rich activation domains have not been observed to function in yeast. As potential partners for TAF110, the Sp1 activation domains were fused to the acidic activation domain of GAL4. Each of the Sp1 glutamine-rich activation domains A or B can independently interact with full-length TAF110 as judged by activation of the reporter gene. In these experiments, yeast bearing an integrated GAL1:lacZ fusion were transformed with two plasmids: (1) fusions to the DNA binding domain of GAL4 (residues 1-147), and (2) fusions to the acidic activation domain (AAD; residues 768-881 of GAL4), and the resulting β-gal activity was measured (expressed in units/ mg of protein). Interestingly, domain A of Sp1 appears to interact more efficiently than domain B, and this correlates well with the previous finding that A is a better activator for transcription than domain B (Courey and

Tjian, 1988). As in Drosophila cells, residues 1-308 of TAF110 are sufficient for the interaction, while regions 1-137 and 138-308 are inactive. The full-length TAF110 fusion is more active than the N308 construct in this assay. Although this effect may be due to differential protein expression, it is possible that the C-terminal regions of TAF110 contribute to interactions with Sp1. The protein-protein interaction assay in yeast further supports the idea that TAF110 interacts, directly or indirectly, with the activation domains of Sp1, and the strength of this interaction appears to be correlated with transcriptional function.

The other TAF proteins that have been tested in the superactivation assay or the yeast assay displayed no detectable interaction with Sp1. However, the GAL4 fusion proteins that these assays rely on might not be able to participate in all the correct interactions because some surfaces could be sterically blocked. Therefore, additional strategies, such as the use of full length Sp1, are used to test for other potential interactions.

dTAF110 does not interact with other activators tested

To determine whether the interaction between Sp1 and TAF110 is specific, or whether other types of activators also interact with TAF110, we used the yeast assay to test a variety of other activation domains including the acidic activation domain of GAL4 (Ma and Ptashne, 1987) and the proline-rich activation domain of CTF (Mermod et al. 1989). Neither of these two activators displayed any interaction with TAF110 in the yeast assay. In addition we tested activation domains from the Drosophila proteins Antennapedia (Antp) and bicoid (bcd), both of which are glutamine-rich. Surprisingly, both of these glutamine-rich domains failed to interact with TAF110 in the yeast assay. Since TAF110 can interact with both Sp1 domains A and B, which have no significant homology other than high glutamine content, but not Antp and bcd which are even more glutamine-rich than Sp1, it appears that glutamine content alone may not be a sufficient criterion for the classification of functionally similar activation domains. In this regard, it may be useful to draw a distinction between the Sp1 activation domains, which are approximately 25% glutamine and flanked by serine/threonine rich sequences, and the bcd and Antp sequences, which are partially composed of uninterrupted stretches of glutamines and lack adjacent serine/threonine sequences.

The N-terminal region of TAF110, containing the glutamine-and serine/threonine-rich sequences, is able to function as a weak activation domain in Drosophila cells, suggesting that this region can interact with a component of the native TFIID complex. To determine whether the N-terminal region of TAF110 is similar to the Sp1 activation domains which can mediate multimerization, we tested for TAF110-TAF110 interactions. We found that the N-terminal region of TAF110 is able to interact with itself as judged by activation of the lacZ reporter gene in the yeast assay (figure 6A). This is another example of functional similarity between the Sp1 activation domains and the N-terminal region of TAF110, which can interact with each other as well themselves.

TBP and other TAFs tested do not interact with Sp1 in yeast

Since Sp1 synergistically activates transcription through multiple sites even though it does not bind cooperatively to DNA, we sought to determine whether Sp1 works via interactions with multiple targets or coactivators. We therefore tested two other members of the TFIID complex, TAF40 and TAF80. Similar to the superactivation assay in Drosophila cells, neither TAF40 or TAF80 displayed any ability to interact with Sp1 under the conditions of the yeast assay. In addition, the conserved C-terminal domain of TBP was tested for Sp1 interaction in yeast but no interaction was observed. We were unable to test full-length dTBP in this assay because it functions as an activator in yeast when fused to the GAL4 DNA binding domain. These results show that the interaction between TAF110 and Sp1 is quite specific, and that TAF80, TAF40, and the conserved region of TBP do not appear to be targets for Sp1.

Since the TFIID complex is also required at promoters that lack a TATA box, one of the TAFs might be required for promoter recognition through the initiator element. In addition to communicating with promoter-selective factors, the TAFs interact with each other, at least one TAF interacts with TBP, and one interacts with RNA polymerase II or one of the basal factors.

Sp1 binds dTAF110 in vitro

The superactivation assay in Schneider cells and the yeast experiments are both indirect assays for protein-protein interactions. Therefore, we also

determined the ability of Sp1 to bind directly to TAF110 in vitro. Biotinylated oligonucleotides containing Sp1 binding sites were coupled to streptavidin-agarose resin. The resin was incubated with Sp1 that had been over-expressed and purified from HeLa cells infected with a vaccinia virus expression vector (Jackson et al., 1990). After allowing Sp1 to bind DNA on the beads, the unbound Sp1 was washed away. Control resin that lacked Sp1 was also prepared and tested in parallel. These resins were incubated in batch with ^{35}S-labeled TAF110 synthesized in vitro in a reticulocyte lysate. After incubation with the labeled protein, the beads were extensively washed and the bound proteins were eluted in two steps with buffer containing 0.2 M KCl followed by 1.0 M KCl. The 1.0 M salt incubation elutes Sp1 from the DNA. The input, unbound supernatant, and eluted fractions were subsequently analyzed by SDS-PAGE and autoradiography. Samples from the binding reaction were also analyzed by silver staining to detect non-specific binding of proteins present in the reticulocyte lysate.

^{35}S-labeled TAF110 synthesized in vitro in a reticulocyte lysate and incubated with streptavidin-agarose beads with or without DNA-bound Sp1. Protein fractions were run on SDS-PAGE and analyzed by autoradiography or by silver staining. After allowing TAF110 to bind Sp1, the beads were pelleted and the supernatant containing the unbound proteins was collected. The resin was washed 4 time. The specifically bound proteins were eluted by incubating the beads in buffer containing 0.2 M KCl, followed by 1.0 M KCl. The Sp1 protein bound to the DNA is eluted by treatment with 1.0 M salt. Labeled TAF110 protein is detectable in the eluted fractions. No detectable TAF110 protein bound to the DNA affinity resin in the absence of Sp1 protein. Quantitation of these results by analysis of the gel in a PhosphorImager (Molecular Dynamics) indicate a 60-fold greater binding by labeled TAF110 to the Sp1-containing resin. The silver stained gel showed that Sp1 is the major species in the eluate indicating that the unlabeled proteins in the extract are not able to bind Sp1.

These data show that TAF110 is selectively retained on the resin containing DNA-bound Sp1, but TAF110 does not bind the control resin that lacks Sp1. Most of the bound TAF110 elutes with the Sp1 at 1.0 M KCl with a lower amount eluting at 0.2 M KCl. Analysis of the fractions by silver staining indicates that Sp1 is the major protein detectable in the high salt eluate, indicating that the

unlabeled proteins present the reticulocyte lysate, which constitute the vast majority of the total protein in the input, are not non-specifically binding to Sp1 in this assay. To rule out the possibility that an intermediary protein, perhaps some other TAF or other eukaryotic protein, was required for the Sp1-TAF110 interaction, this experiment was repeated using ^{35}S-labeled TAF110 synthesized in an in vitro transcription/translation extract derived from E. coli (Skelly et al., 1987). The TAF110 protein synthesized in the prokaryotic system was also specifically retained on the Sp1 affinity resin providing further evidence that Sp1 can bind directly to TAF110.

As an additional test of specificity, we also determined if deletion mutants of TAF110 could bind to Sp1 in this in vitro assay (mutants are expressed from the N-terminal). A 1-137 mutant was not able to bind Sp1 in vitro, while some binding was obtained with a 1-308 mutant. Mutants of 308-921, 447-921, and 571-921 were all effective in binding Sp1, while C-termina deletions beyond 852 from these mutants eliminated Sp1 binding. These results indicate the importance of a 852-921 region and a 137-308 region of TAF110 in transcription activator interaction.

TAF110 does not directly bind TBP

Our experiments indicate that TAF110 cannot directly bind to TBP by itself and that at least one additional TAF is required to connect TAF110 and TBP. For example, α-TAF110 antibodies fail to coprecipitate both in vitro expressed TAF110 and TBP and similarly with α-TBP antibody.

Exemplary Experimental Procedures

Purification of the TFIID complex

For the in vitro transcription assay, the TFIID complex was immunopurified from the partially purified TFIID fraction (Q-sepharose fraction, 0.3 M KCl eluate) (Dynlacht et al., 1991) using the α-dTBP monoclonal antibody 42A11 coupled to protein G-sepharose (Pharmacia). The immunoprecipitates were washed with 0.1 M KCl-HEMG-ND buffer (25 mM HEPES pH 7.6, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.1 mM DTT) and the TFIID complex was eluted from the antibody by addition 10 mg/ml of the peptide

mimicking the epitope of 42A11 (sequence: NH2-RPSTPMTPATPGSADPG-COOH) in HEMG buffer containing 0.5 M guanidine-HCl. The eluate was dialyzed against 0.1 M KCl-HEMG-ND, and then assayed for transcription activity.

Purification of dTAF110

Nuclear extracts derived from approximately 1 kg of Drosophila embryos were prepared and fractionated as previously described (Dynlacht et al., 1991; Wampler et al., 1990). For protein sequencing, the TFIID complex was purified with polyclonal α-dTBP antibodies as previously described (Dynlacht et al., 1991) or with a monoclonal antibody as described above. The TAFs were separated from TBP by elution of the protein A-antibody resin with 0.1 M KCl-HEMG buffer containing 1.5 mM DTT, 0.1 % LDAO (lauryl dimethylamineoxide), and 1M Gd-HCl. The TAFs were eluted by batch incubation of the resin with an equal volume of buffer for 25 min at 4 °C. This procedure was repeated and the two supernatants were combined. Urea was added to 8 M, DTT to 10 mM, and cysteines were modified with 4-vinylpyridine.

Two approaches were used to separate the TAFs: HPLC and PAGE. Under the HPLC approach, the TAFs were fractionated by reverse phase HPLC on a 300 angstrom C4 column (2.1 X 30 mm). The proteins were eluted with a gradient from 20-70% buffer B (buffer A = 0.1% TFA, 1% n-propanol; buffer B = 0.1% TFA, 1% n-propanol, 60% isopropanol, 30% acetonitrile). TAF110 consistently eluted at 35% buffer B. Fractions containing TAF110 (approximately 5 μg) were lyophilized, resuspended in 100 mM TRIS, pH 8.0, and 2 M urea, and incubated at 55 °C for 10 min. 150 ng of the protease lys C was added and the protein was digested for 20 hr at 37 °C. Peptides were chromatographed and sequenced as previously described (Williams et al., 1988).

Under the gel electrophoresis approach, the TAFs were separated by electrophoesis and transferred to membranes. The separated TAFs were digested with LysC or trypsin and the resultant peptides eluted, chromatographed and sequenced. See Fernandez et al., (1992) Analytical Biochemistry 201, 255-264.

In vitro transcription assay

Transcription factor fractions were reconstituted with basal factor fractions derived from 0-12 hr Drosophila embryo nuclear extracts essentially as previously described (Dynlacht et al., 1991) except that TFIIB was separated from TFIIE/TFIIF and pol II was fractionated further on a phosphocellulose column. Each reaction contained 0.5 ug of the TFIIB fraction (S-sepharose 0.5 M eluate), 1.5 ug of the TFIIE/TFIIF fraction (S-sepharose 0.25 M eluate), and 0.25 mg of the pol II fraction (phosphocellulose 0.4 M eluate). Some reactions contained 1.5 ug of the TFIID fraction or 2 ng of purified, recombinant dTBP that had been expressed in E. coli (Hoey et al., 1990). The template for the in vitro transcription reaction was BCAT (Lillie and Green, 1989) containing 3 Sp1 binding sites, and transcription was assayed by primer extension.

Generation of antibodies against the TAFs

Immunopurified TFIID complex (approximately 10 ug/ injection) was mixed with Ribi's adjuvant and injected intraperitoneally into a Swiss-Webster mouse at days 0, 7, and 21. The initial immune response was monitored at day 28 and boosted further by two biweekly injections of more antigen. After an intravenous injection of one further dose of antigen the spleen was dissected out and electrofused with myeloma cells. Approximately 600 supernatants from 96-well dishes (each well containing on average 5 independent hybridomas) were assayed on western strip blots for cross-reactivity with immunopurified TFIID complex proteins. Hybridomas from wells producing anti-TAF and/or anti-TBP antibodies were cloned by limited dilution and tested by Western blotting and immunoprecipitation assays.

Cloning of TAF110 cDNAs

The polyclonal antiserum obtained from the immunization scheme described above was used at a 1/1000 dilution to screen approximately 5×10^5 plaques of a size-selected (> 1.8 kb) 9-12hr lgt11 Drosophila cDNA library (Zinn et al., 1988). Positive clones were plaque-purified to homogeneity and tested for cross-reactivity against anti-TAF monoclonal antibodies of known specificity. One clone, λ106, cross-reacted strongly with several independent anti-TAF110 hybridomas.

Insert DNA (2.6 kb) from λ106 was purified and labeled using Klenow polymerase and random hexamer priming (Amersham). 10^6 recombinant phage from a cDNA library (Poole, et al., 1985) prepared from 3-9 hour Drosophila embryos were screened as previously described (Kadonaga et al., 1987). 24 positives were obtained in duplicate on the primary screen; 12 of these were randomly selected for rescreening, and 10 of 12 were positive on the secondary screen. All 10 of these cDNA clones were found to be related to each other on the basis of restriction mapping and cross-hybridization. The largest cDNA clone of 4.6 kb, λ110-5, was completely sequenced, and two other clones of 3.1 kb, λ110-1, and 2.1 kb, λ110-2, were partially sequenced. The inserts were subcloned into pBS-SK (Stratagene) in both orientations, a nested set of deletions was constructed with exonuclease III, and the clones were sequenced by the dideoxy method. The λ110-1 clone was found to be 37 nucleotides longer at the 5' end than the λ110-5 clone and missing 1.5 kb on the 3' end. The SEQUENCE ID NO: 1 is a composite of the λ110-1 and λ110-5 sequence.

Expression of dTAF110 protein

An NdeI site was created at the initiating methionine using a PCR based strategy. A 3.1 kb NdeI-BssHII fragment containing the entire coding sequence was subcloned into the SmaI site of the baculovirus expression vector pVL1392 (Pharmingen). Recombinant baculoviruses were selected by co-transfection of Sf9 cells with the expression vector and linear viral DNA as described by the supplier (Pharmingen). Samples for the western blot were prepared by infecting SF9 cells with recombinant virus obtained from the transfection supernatant. Three days after infection the cells were harvested, washed, resuspended in HEMG buffer, and lysed by sonication. The protein concentration was measured by Bradford assay. After electrophoresis proteins were transferred to nitrocellulose: TAF110 protein was detected using the monoclonal antibody 3E7.

Transfections

Transfection of Schneider cells (line SL2) was carried out as previously described (Courey and Tjian, 1988) except that the transfections were performed in 60 mm dishes. The expression vector for all proteins used in this study was pPac,

which contains the Drosophila actin 5c promoter. TAF110 sequences were fused in frame to GAL4 DNA binding domain, residues 1-147. The following restriction fragments of the TAF110 cDNA were used: N137, NdeI-ClaI; N308, NdeI-SalI, 138-308, ClaI-SalI; 87-308, HincII. The constructs were checked by sequencing across the fusion junctions. The amounts of DNA used were as follows: 100 ng of the pPacGAL4 derivatives, 500 ng of the pPacSp1N539, and 2.5 ug of the reporter gene pG5BCAT (Lillie and Green, 1989). CAT assays were performed and quantitated as previously described (Courey and Tjian, 1988).

Yeast Methods

The yeast strain Y153 (a, gal4, gal80, his3, trp1-901, ade2-101, ura3-52, leu2-3, 112, URA3::Gal1:lacZ, LYS2::Gal-His3) was transformed with two plasmids according to the method of Shiestl and Gietz (Schiestl and Gietz, 1989). The Gal4 DNA binding domain hybrids were constructed in the vector pAS1. pAS1 is a 2μ plasmid with TRP selection that expresses fusions to Gal4(1-147) from the ADH promoter. For expression of GAL4(1-147), an XbaI linker containing stop codons in all three reading frames was inserted in pAS1 immediately downstream of the GAL4(1-147) coding sequence. G4-110 (fl) contains the entire coding region of the TAF110 on an NdeI-BssHII fragment, and the shorter G4-110 fusions contain fragments as described for the Drosophila expression vectors. G4-80 (fl) contains an NdeI-XbaI fragment that includes the entire coding region of Drosophila TAF80. G4-40 (fl) contains an NdeI-EcoRV fragment encoding Drosophila TAF40. G4-dTBP(191C) contains an NdeI fragment derived from pAR-191C containing the conserved C-terminal domain (Hoey et al., 1990). The reading frame across all fusion junctions was verified by sequencing, and the protein expression was verified by western blot analyses with either α-TAF or α-GAL4 antibodies, with the exception of G4-110(N137).

The acidic activation domain fusions were constructed in the vectors pGAD1F, pGAD2F or pGAD3F which differ only in the reading frame of a unique Bam site (Chien et al., 1991). These 2μ plasmids with LEU2 selection express fusions to activating region II (residues 768-881) of GAL4 from the ADH promoter. Sp1 region A consists of amino acids 83-262 and Sp1 region B consists of residues 263-542; these were cloned as BamHI-BglII fragments from the

plasmids pKSABg10 and pKSBG respectively. The C-terminal 100 amino acids of CTF1 (residues 399-499) were cloned as a BglII-EcoRI fragment (Mermod et al., 1989). The Antp construct was made by subcloning a BamHI fragment containing the activation domain (Courey et al., 1989). Bcd residues 249-489 (Driever et al., 1989) were cloned on a SalI fragment derived from pPac-bcd. The reading frame across all fusion junctions was verified by sequencing.

Transformed yeast were assayed qualitatively after growth on media containing X-gal. Quantitative ß-galactosidase assays were performed as described (Himmelfarb et al., 1990) except cells were grown to mid log in selective media containing 2% glucose. Assays were performed in triplicate and activity is expressed as units/mg of total protein.

In vitro protein-protein interaction assay

A 3.1 kb NdeI-BssHII fragment containing the entire TAF110 coding region was subcloned into the plasmid pTbSTOP (Jantzen et al., 1992), which contains the b-globin untranslated leader downstream of a T7 promoter. The plasmid was linearized with XbaI, and the gene was transcribed in vitro with T7 RNA polymerase. ^{35}S-met labeled protein was synthesized in vitro in a rabbit reticulocyte lysate (Promega). Alternatively, TAF110 was synthesized in vitro in an E. coli derived S30 transcription/translation extract (Skelly et al., 1987). Sp1 protein was overexpressed in HeLa cells using a vaccinia virus expression vector (Jackson et al., 1990) and purified by wheat germ agglutinin (WGA) affinity chromatography (Jackson and Tjian 1990), prior to DNA affinity purification as outlined below.

DNA affinity resin was prepared as follows: 5'-biotinylated oligonucleotides containing 4 Sp1 binding sites, GCA(AGGGGCGGGGCT)$_4$T and its complement, were annealed and coupled to streptavidin-agarose beads (Pierce) by incubating overnight at room temperature. The beads were incubated with WGA-purified Sp1 in buffer Z' (25 mM HEPES, pH 7.6, 20% glycerol, 0.1% NP-40, 10 mM $ZnSO_4$, 1 mM DTT) containing 0.1 M KCl for 2 hours at 4 °C. Sp1 was bound to the resin at a concentration of approximately 1 mg/ml of beads. ^{35}S-labeled TAF110 was incubated in batch with 15 ml of the DNA affinity resin in Z'+ 50 mM KCl, with or without Sp1, for 4 hours at 4 °C. The beads were

washed 4 times with 1 ml of the same buffer, and eluted with Z' + 0.2 M KCl, followed by Z' + 1.0 M KCl. The eluted proteins were TCA-precipitated and analyzed by SDS-PAGE. Before autoradiography, the gel was fixed and treated with Amplify (Amersham).

Detection of Direct TBP/TAF Interactions on Protein Blots

Immunopurification of the Drosophila TFIID complex using anti-TBP antibodies results in the purification of a large multiprotein complex consisting of TBP and 7 major TAFs. To identify TAFs which can bind directly to TBP we probed a blot containing renatured TAFs with a 32P-labelled TBP-GST fusion protein. After washing off unbound TBP-fusion protein and exposing the blot to X-ray film a strong signal was seen which coincided with the position of dTAFII-250K on the gel. Further experiments revealed that a truncated version of TBP, consisting of the highly conserved C-terminal domain, is sufficient to mediate this interaction. We also tested other fractions containing basal factors (Wampler et al., 1990; Dynlacht et al., 1991), including TFIIB, E/F and RNA polymerase II, and failed to detect specific signals. We conclude that TBP and TAFII-250K interact directly and that TAFII-250K is present in the TFIID fraction but not associated with TFIIB, E, F or RNA polymerase II.

Molecular Cloning and Characterization of the dTAFII-250K Gene

Having identified dTAFII-250K as a candidate for a direct TBP-TAF interaction we decided to clone the corresponding gene. The low abundance and large size of dTAF(II)-250K disfavours cloning strategies based on protein microsequencing. Instead, we were able to obtain monoclonal antibodies which specifically (and exclusively) recognize dTAF(II)-250K on Western blots. To show that dTAF(II)-250K is indeed a genuine component of the TFIID complex, we used two of these monoclonal antibodies, 2B2 and 30H9, to carry out immunoprecipitations from the TFIID fraction. The pattern and stoichiometry of TAFs and TBP is indistinguishable from the ones described previously using either anti-TBP (Dynlacht et al., 1991) or anti- dTAF(II)-110K (Hoey et al., 1993) antibodies. We cloned the gene encoding the Drosophila dTAF(II)-250K by screening a lgt11 expression library prepared from 6-12 hour old embryos (Zinn et

al., 1988) with hybridoma supernatants containing either 2B2 and 30H9 anti-dTAF(II)-250K monoclonal antibodies. Five partial cDNA clones were obtained, which all cross-hybridized with each other at high stringency. Restriction mapping and sequence analysis confirmed that they were indeed derived from the same gene. Two of these cDNAs, ID-1 and ID-2, allowed us to establish a composite open reading frame spanning 4.5 kb (fig. 2). Attempts to isolate additional cDNA clones encoding N-terminal regions of dTAF(II)250 or 5'-RACE experiments have so far been unsuccessful. Genomic DNA sequencing allowed us to extend the open reading frame by approximately 1 kb before encountering noncoding (presumanbly intronic) sequences. Inspection of the open reading frame encoded by the cDNA clones reveals a protein sequence which displays an extensive similarity to the human 'Cell Cycle Gene 1' (CCG1) gene previously described by Sekiguchi et al., 1991. Many of the sequence elements defined in the CCG1 genes are also present in the dTAF-250K encoding sequence. Interestingly, however, we detected a 35 amino acid insertion in the region which Sekiguchi et al. putatively identified as an HMG box. This insertion causes substantial disruption of the spatial alignment with the consensus sequence. We also used the ID-2 cDNA fragment to map the dTAFII-250K gene to position 32E1-2 (left arm of chromosome II) by in situ hybridization. This location does not contain any previously characterized genes and currently no deletions spanning that regions are available. Since dTAF-250 seems to be present in all or the majority of the TFIID complexes present within cells and seems to provide essential contact points with TBP and TAFs (see below) we expect that a deletion of the 32E1-2 locus would cause a lethal phenotype.

Expression of the C-terminal domain of dTAF(II)-250K in Insect Cells

To study the functional properties of the proteins encoded by these cDNAs we decided to express the protein encoded by the reading frame of our longest cDNA, ID-1. Because of the expected large size of the protein encoded we chose the baculovirus system. After subcloning of the fragment into expression vector pVL1393 and transfecting the construct into Sf9 cells we detected expression of a 180K protein (subsequently referred to as DN250) which cross-reacted strongly with several anti-TAF250 monoclonal antibodies recognizing a variety of epitopes

in different parts of the 250K TAF. We detected no cross-reactivity between our antibodies and any endogenous Spodoptera TAF250 homologs which might be present in Sf9 cells.

The C-terminal Domain of the dTAF(II)250K Is Sufficient for TBP Binding

To study whether DN250 was capable of interacting with TBP we immunopurified the protein from infected cells. Monoclonal antibody 30H9 was bound to protein A or G beads and incubated with extracts from baculovirus infected cells. Under these conditions DN250 is specifically immobilized on the beads. After washing off unbound material we added an extract containing partially purified TBP (also expressed in the baculovirus system). TBP was specifically bound to beads carrying the immunopurified TAF250-C180 protein whereas beads containing antibody only failed to do so. Further evidence for this direct TBP-TAF interaction by carrying out protein blots. The ability of a protein representing appr. 60% of the full-length 250K protein to bind TBP demonstrates conclusively that the cloned C-terminal part is sufficient for TBP binding.

Gelshift Analysis of the DN250/TBP Complex

TBP is the only component of the general transcriptional machinery capable of sequence-specific binding to the TATA box. We therefore were interested to see how interaction of TBP with TAF250-C180 affected the specificity and affinity of DNA binding. TBP was added to a 32-P labelled DNA fragment containing the -33 to +55 region of the adenovirus major late promoter and DNA-binding was monitored using a gelshift assay. The intensity of probe DNA shifted by TBP increased substantially in presence of purified TAF250-C180 wheras TAF250-C180 alone did not detectably bind to DNA. To investigate whether this enhanced affinity of the TBP/TAF250-C180 complex for DNA was due to additional contacts with DNA provided by the TAF250-C180 protein we carried out footprinting studies, again using the adenovirus major later promoter region as a probe.

TAF250 and TAF110 Specifically Interact With Each Other, even in Absence of TBP

Since we have not observed any of the cloned Drosophila TAFs to bind to TBP we investigated whether they would interact with the TBP/ d250KdeltaC180 complex. 35S-labelled 110K protein (Hoey et al., 1993) was synthesized in an in-vitro translation system and incubated with TAF250-C180 protein in presence and absence of TBP. As shown in fig. 5 we found that the 110K TAF binds specifically to dTAF(II)250K-C180 in the presence and absence of TBP thus indicating that the two proteins bind independently to two distinct domains within the 250K TAF. The affinity and specificity of this interaction is sufficiently high to allow selective purification of TAF110 from a crude baculovirus extract expressing the recombinant protein by using TAF250-C180 immobilized on beads.

Protein Blot Analysis

pGEX-2TK was linearized with SmaI, phosphatase-treated and the ligated with gel-purified NdeI fragments of either pARdTFIID or pARdTFIID-191C (Hoey et al., 1990). Generation of 32-P labelled GST fusion protein, protein blotting and hybridization were carried out essentially as described in Kaelin et al., 1992.

Generation of anti-dTAFII-250K Hybridoma Cell Lines

The monoclonal antibodies described in this study were derived as described in Hoey et al.(1993). Briefly, a Swiss-Webster mouse was immunized with intact immunopurified Drosophila TFIID complex. After fusion hybridoma supernatants containing anti-dTAFII-250K antibodies were selected using stripblots containing SDS-gel-separated TBP and TAFs. Two such cell lines, 2B2 and 30H9, were then cloned to homogeneity by limited dilution.

Isolation of dTAFII-250K cDNA and Genomic Clones

Approximately 5x105 independent plaques of a size-selected (>= 1.8kb) Drosophila lgt11 library prepared from Drosophila embryos (Zinn et al., 19..) were screened with two independent anti-dTAFII-250K monoclonal antibodies, 2B2 and 30H9. All the positives identified cross-hybridized at high stringency with each other on the DNA level. Restriction mapping and sequence analysis showed that all

of the clones were derived from the same gene. cDNA clones ID1 and ID2 contained inserts of 1.5 and 4.0 kb, respectively, and were sequenced to completion. ID2 was found to extend 500 bp further towards the 5' end of the gene and was used to isolate genomic clones IDASH3 and IDASH4 (Sau3A partially digested DNA cloned into IDASH).

Sequencing Strategy

We employed the gd transposon-directed sequencing strategy (Gold Biosystems) as described in Strathmann et al., 1991. DNA fragments of interest were subcloned into the plasmid vector pMOB1 and electroporated into DPWC cells. After conjugation with the recipient host BW26 the mixture was plated out on kanamycin/carbenicillin plates. Transposon insertion points were mapped by PCR. Clones with the desired transposon locations were then grown up and sequenced using transposon-specific primers with 35S-dATP or the Pharmacia A.L.F. Sequencer.

Expression of a Truncated Version of dTAFII-250K (DN250) in the Baculovirus System

cDNA #5 was inserted into the EcoRI site of Baculovirus-expression plasmid pVL1393 (Pharmingen). The resulting construct was co-transfected with 'BaculoGold' viral DNA (Pharmingen) into Sf9 cells. After 3 days cells were harvested and expression of the DN250 protein was monitored by Western blotting using the anti-dTAFII-250K monoclonal antibody 2B2. The recombinant virus-containing supernatant was used to infect large scale cultures of Sf9 cells. We typically prepared whole cell extracts from 1 liter of plate cultures of infected Sf9-cells by sonicating them in HEMG-ND/0.1 M KCl (HEMG-ND contains 25mM HEPES, pH7.6, mM MgCl2 0.1 mM EDTA, 0.1 % NP40, 1 mM PMSF, 1.5 mM DTT, 5mg/ml leupeptin). The supernatant was partially purified (approximately 5 fold) by chromatography over Q-sepharose (Pharmacia) with step gradient elution (HEMG containing 0.1 M, 0.2, 0.4 and 1.0 M KCl, respectively). dTAFII-250K(C180) eluted in the 0.4 M step('Q.4' fraction). After dialysis against HEMG-0.1M KCl the extract was frozen in aliquots and used for the immunopurification/coprecipitation studies.

Coimmunopreciptiation Studies

Protein G-beads were preloabded with monoclonal antibodies and incubated with various cell extracts from Baculovirus-infected cell fractions or 35S-labelled dTAFII110 prepared by in vitro translation. After 45 minutes on ice, unbound protein was removed with several washes with HEMG-ND.

hTAFII250 purification and cloning

We previously reported the isolation of hTFIID by affinity chromatography using antibodies specific to TBP. The purified complex contains at least seven distinct TAFs ranging in molecular weight from 30-250 kD which copurify with TBP. We were particularly interested in characterizing the 250 kD species because this subunit of TFIID appears to bind TBP directly as determined by Far Western analysis. Using affinity-purified TAFs to immunize mice, we generated both polyclonal and monoclonal antibodies that crossreact with different TAFs. We used these antibodies to screen lgt11 expression cDNA libraries and several clones were isolated, including IH1 which contains a 1.1 kb insert. To determine which, if any, TAF is encoded by IH1, we expressed this cDNA as a GST fusion protein, purified the tagged protein by glutathione affinity chromatography, and raised antibodies against this recombinant protein. Antisera directed against GST-IH1 specifically crossreacted with the 250 kD TAF, indicating that a portion of the gene encoding hTAFII250 had been isolated.

Next, we determined the DNA sequence of IH1 and discovered that this open reading frame is related to the previously identified human gene, CCG1, which had been implicated in cell cycle regulation. Specifically, a temperature-sensitive mutant hamster cell line, ts13, is arrested at G1 a few hours before entering S phase at the non-permissive temperature. Expression of human CCG1 in ts13 overcomes this cell cycle block. Since IH1 only encoded a small portion of hTAFII250, we isolated several additional clones from a primary HeLa cDNA library, including IH2, which contained a 5.3 kb insert. The construction of a full-length hTAFII250 cDNA revealed the predominant hTAFII250 RNA species characterized in HeLa cells encodes 21 additional amino acids between residues 177 and 178 relative to CCG1. Interestingly, we sequenced several other cDNAs containing internal insertions or deletions when compared to CCG1. This

finding suggests that multiple hTAFII250-related proteins may be generated by alternate splicing of a primary transcript.

Although the finding that a cDNA isolated by antibodies directed against TAFs encodes a cell cycle gene is exciting, it was important to provide some functional evidence that this clone indeed encodes a bona fide TAF which is a subunit of TFIID. We first asked whether the recombinant hTAFII250 expressed in a vaccinia virus system becomes associated with the endogenous TFIID complex in HeLa cells. To distinguish between the recombinant and endogenous protein, we engineered a version containing a hemagglutinin antigen (HA) epitope at the N-terminus of hTAFII250. Antibodies against TBP were used to immunopurify the TFIID complex from HeLa cells infected with either recombinant or control vaccinia virus. The immunopurified complexes were subjected to gel electrophoresis and analyzed by Western blot analysis using either a monoclonal anti-HA antibody to detect the HA-tagged molecule or monoclonal antibody 6B3, raised against the endogenous hTAFII250. The anti-HA antibody crossreacted specifically with a 250 kD protein only in the TFIID complex prepared from recombinant hTAFII250 virus infected HeLa cells but not control infected cells. As expected, 6B3 recognized both the recombinant hTAFII250 and the endogenous protein. Thus, we conclude that the recombinant hTAFII250 associates with TBP in vivo and is part of the TFIID complex.

To test for a direct interaction between hTAFII250 and TBP, we performed a Far Western analysis with radiolabelled TBP and antibody immunopurified HA-tagged hTAFII250. The full-length hTAFII250 is capable of interacting directly with TBP in vitro, even in the absence of other TAFs or coactivators. These results and the analysis of the independently cloned Drosophila TAFII250 suggest that this largest TAF is responsible for the initial assembly of the TFIID complex by binding directly to TBP and other TAFs.

The important role of hTAFII250 in the formation of a TFIID complex prompted us to define more precisely its interaction with TBP. For these studies we employed the two hybrid system carried out in yeast cells. Using this rapid and convenient assay for protein:protein interactions, we observed that a hybrid construct containing hTAFII250 fused to the DNA binding domain of GAL4, G4(1-147), interacted selectively and efficiently with human TBP attached to the

acidic activation domain of GAL4, G4(768-881). Yeast expressing both of these proteins produced high levels of b-galactosidase due to increased transcription of a lacZ reporter construct, containing GAL4 binding sites. Interestingly, hTAFII250 also interacts efficiently with a truncated version of human TBP which contains only the conserved C-terminal 180 amino acids. By contrast, a construct containing the "species-specific" N-terminal domain of human TBP failed to interact with hTAFII250. These results are in agreement with Far Western experiments using radiolabelled cTBP and nTBP as probes and suggest that residues 160 to 339 on the outer surface of TBP may be responsible for hTAFII250 binding.

Our unexpected finding that hTAFII250 is related to CCG1 suggests a rather intriguing link between a subunit of TFIID and expression of genes involved in cell cycle control. Interestingly, CCG1 is a nuclear phosphoprotein with several domains characteristic of transcription factors including a putative HMG-box and a proline-rich cluster. Based on these structural motifs, Sekiguchi et al. suggested that CCG1 might work as a sequence-specific transcription factor needed for regulating genes involved in the progression through G1. However, it now seems clear that CCG1 or a related product is part of the TFIID complex and is not a promoter-specific transcription factor. Therefore, it seems more likely that the G1 arrest in ts13 is due to the failure of a defective TFIID complex to mediate activation by a subset of cellular transcription factors that govern cell cycle genes, e.g. thymidine kinase and dihydrofolate reductase genes. The presence of a putative DNA binding domain, the HMG box, may suggest that once hTAFII250 forms a complex with TBP, some portion of this large subunit of TFIID may contact DNA, perhaps downstream of the initiation site.

Immunoaffinity purified hTFIID complex: Interaction with hTBP and production of hTAFs-specific antibodies

A. Immunoprecipitation reactions were carried out according to a modified version of previously described procedures (Tanese et. al.). 0.5 mg of affinity purified a-hTBP antibody was added to 200 mg of hTFIID (phosphocellulose 0.48 - 1.0 M KCl) fraction, and the mixture nutated for 2 - 4 hrs at 4oC. Protein A Sepharose was then added and nutation continued for an additional 2 - 4 hrs.

Antibody-antigen complexes were pelleted by low-speed centrifugation, washed four times with 0.1 M KCl - HEMG (25mM Hepes, 12.5 mM MgCl2, 0.1 mM EDTA, 10% glycerol) containing 0.1% NP-40 and 1mM DTT. The immunoprecipitated hTFIID complex was subjected to 8% SDS-PAGE and silver stained. For Far Western analysis, the proteins were blotted onto nitrocellulose membrane and hybridized with 35S-labeled hTBP (Kaelin et al.). pTbhTBP was used to in vitro transcribe hTBP RNA which was in vitro translated using 120 mCi 35S-methionine (> 1000 Ci/mMol, Amersham) in reticulocyte lysate (Promega).

B. Antigen used to immunize mice for antibody production was prepared as follows. The immunoprecipitated hTFIID complex, purified from 250 litres of HeLa cells, was eluted from the Protein A Sepharose - antibody complex with 0.1 M KCl - HEMG containing 1 M guanidine - HCl, 0.1% NP-40, and 1 mM DTT. Under these conditions TBP remained bound to the antibody. The eluted TAFs were dialized against 0.1 M KCl - HEMG containing 0.1% NP-40 and 1 mM DTT. The mixture of proteins containing 1-2 mg of each TAF was used to immunize a mouse. Test bleeds were taken and the immune response monitored by Western blot analysis. After a series of five boosts, the mouse was sacrificed and the spleen was used for the production of monoclonal antibody producing hybridoma cells lines. The identification of hybridoma cell lines producing hTAF specific antibodies was determined by Western blot analysis of eluted TAFs.

Cloning and identification of the 250 kD subunit of hTFIID complex as CCG1

A. An expression screen of 2.4 x 106 PFU from a lgt11 HeLa S3 cDNA library (Clontech) was carried out using the a-hTAFs polyclonal serum described above. 38 primary signals were identified of which 6 were plaque purified. 1 phage DNA was prepared and analyzed by EcoRI restriction enzyme digestion. IH1 contained a 1.1 kb insert which was subcloned into the EcoRI site of pGEX1 (Pharmacia) to express a GST-IH1 fusion protein. The resulting construct was transformed into Escherichia coli TG2, and following induction with 0.5 mM IPTG, the induced protein was purified on glutathione Sepharose 4B beads (Pharmacia). 2 mg (per injection) of the fusion protein was used to immunize a mouse. Test bleeds were taken and used for Western blot analyses.

B. Poly(A)+ RNA from HeLa cells was used for construction of a directional cDNA library in lZAPII (Stratagene) as described previously (Ruppert et al. 1992). Using a randomly 32P-labeled probe derived from the IH1 cDNA insert, 15 independent cDNA clones were isolated from 1.2 x 106 original PFU. The cDNA inserts were rescued by the zapping procedure (Short et al.) and characterized extensively by restriction enzyme analysis and Southern blotting. The longest cDNA clone isolated from IH2 contains a 5.3 kb insert, revealing an extended 3' untranslated region but missing about 1.15 kb of 5' sequences when compared to CCG1. This 5' region was generated by PCR using conditions described previously (Ruppert et al.). Two set of PCR primers were designed according to the CCG1 cDNA sequence (Sekiguchi et al). PCR-I, forward primer #1: 5'-TATTTCCGGCATATGGGACCCGGCTG-3' (position 40 to 65, containing an engineered NdeI restriction site at the translation start codon) and reverse primer #2: 5'-GAAGTCCACTTTCTCACCAG-3' (position 578 to 597). PCR-II, forward primer #3: 5'-TACCAGCAGCATATGGGGAGCTTGCAG-3' (position 421 to 447) and reverse primer #4: 5'-GCTCTAAGGAAGCCAGCCTGCCAGGCTTG-3' (position 1343 to 1371). All PCR products were subcloned into pBluescript KS (Stratagene) and sequenced. The most abundant product of PCR-II, a 1 kb fragment, included a 63 bp in frame insertion ,while a minor 330 bp fragment revealed a 618 bp in frame deletion with respect to the CCG1 cDNA. To generate a full-length hTAFII250 cDNA, the product of PCR-I and the 1 kb PCR-II product were joined via the shared SmaI restriction site. Subsequently the 1.2 kb XbaI fragment of the resulting plasmid was cloned into XbaI cut pH2 to generate the full-length cDNA clone phTAFII250.

Analyses of hTAFII250 and hTBP interaction

A. To construct an HA-tagged version of hTAFII250 we generated a plasmid, pSK-HAX, containing the hemagglutinin antigen (HA) epitope, factor X cleavage site, and in frame NdeI cloning site. A 6.3 kb NdeI/Asp718 fragment from phTAFII250 was inserted into pSK-HAX to generate pHAX-hTAFII250. A 6.0 kb SpeI fragment thereof containing the complete coding region of hTAFII250,was inserted into the XbaI site of the vaccinia virus expression vector pAbT4537 (Applied bioTechnology Inc.). Extracts from recombinant virus,

vhTAFII250, or control virus (New York City Board of Health strain of vaccinia virus) infected HeLa cells (Dynlacht 1989) were fractionated by phosphocellulose chromatography as described (Tanese et al.). hTFIID complexes from the 0.48 - 1.0 M KCl fraction were immunoprecipitated with affinity-purified a-hTBP antibodies, subjected to 8% SDS-PAGE and analyzed by Western blotting.

B. To generate an HA-tagged version of hTAFII250 in the baculovirus expression system, we first generated new baculovirus vectors, pVL1392HAX and pVL1393HAX, derived from pVL1392 and pVL1393 (Pharmingen), respectively. These vectors contain the HA antigen epitope, factor X cleavage site, and unique in frame NcoI and NdeI restriction sites. A 6.0 kb NdeI/SpeI fragment from phTAFII250 was inserted into pVL1392HAX creating pbHAX-hTAFII250. Whole cell extracts from either SF9 cells or SF9 cells infected with recombinant baculovirus were prepared in 0.4 M KCl - HEMG (including 0.04% NP-40, 1 mM DTT, 0.2 mM AEBSF, 0.1 mM NaMBS) and used directly for immunoprecipitation with the a-HA antibody. The precipitate was subjected to 8% SDS-PAGE and blotted onto nitrocellulose membrane. The filter was probed first with 35S-labeled hTBP, and subsequently with the monoclonal antibody 6B3.

hTAFII250 interacts with hTBP in yeast

hTAFII250, fused to the DNA binding domain of GAL4 (residues 1-147), was constructed by inserting a 6.0 kb NdeI/BamHI fragment derived from pvhTAFII250 into the pASI vector. The activation domain fusions were obtained by cloning inserts into the pGAD1F vector (Chien et al.). The hybird proteins generated included the acidic activation domain of GAL4 (residues 768-881) fused to either full-length , residues 160-339, or residues 1-159 of hTBP. The above described constructs were transformed into the yeast strain Y153 (a, gal4, gal80, his3, trp1-901, ade2-101, ura3-52, leu2-3, 112, URA3::Gal1:lacZ, LYS2::Gal-His3; as described (Chien et al.) and b-galactosidase assays performed according to published procedures (Hoey et al).

Drosophila TBP and dTAFII250 interact with the C-terminal portion of dTAFII150

Radiolabeled in vitro translated dTAFII150 bound efficiently to immobilized HA-dTBP or dTAFII250ΔN (see Weinzierl et al (1993) Nature 362, 511-517). In

contrast, dTAFII110 and other TAFs failed to interact selectively with dTAFII150, showing that dTAFII150 interacts with at least two subunits of the TFIID complex, dTBP and dTAFII250, which also contact each other.

We also carried out in vivo experiments in which insect Sf9 cells were co-infected with two recombinant baculoviruses, one expressing dTAFII150 and the second expressing either TBP or one of the other TAFs. Complexes were subsequently immunopurified from cellular lysates and analyses by SDS PAGE followed by immunoblotting using antibodies directed against dTAFII150. Coinfection of virus expressing dTAFII150 and either HA-dTBP or dTAFII250ΔN resulted in efficient formation and copurification of heteromeric complexes. Similarly, full-length hTAFII250 bound efficiently to dTAFII150.

Radiolabeled in vitro translated C-terminal 369 residue portion (dTAFII150ΔN) of this protein binds TBP and dTAFII250ΔN with the same effenciency as the full length protein. No significant binding of a N-terminal 786 residue portion (dTAFII150ΔC) was observed: i.e. the interaction interfaces from these proteins are located in the C-terminal portion of dTAFII150.

TSM-1 associates with TBP and TAFII250

Like dTAFII150, TSM1ΔN (C-terminal 920 residue portion) bound efficiently to yTBP as well as HA-dTBP; hence we conclude that yeast contain a TAFII250 and TSM-1 is a TAF.

The activation domain of the Drosophila regulator NTF-1 (Neurogenic Element Binding Transcription Factor-1) interacts with dTAFII150.

NTF-1 immuno-copurifies with dTFIID using anti-dTBP, indicatin that one or more subunits of the dTFIID interacts directly with NTF-1. Using coimmunoprecipitation experiments: dTAFII150 was immunopurified from Sf9 extracts containing dTAFII150, the immobilized TAF was mixed with recombinant NTF-1, the isolated complex was analyzed by SDS-PAGE, and the presence of NTF-1 was detected by protein immunoblot anaysis, showing that NTF-1 directly interacts with dTAFII150.

Next we used a GST-NTF-1 fusion protein containing the N-terminal 284 amino acids of NTF-1 to bind various truncated bersions of dTAFII150, showing that the N-terminal, but not the C-terminal region of dTAFII150 bound to the N-

terminal extended activation domain of NTF-1. Neither dTAFII80 nor dTAFII40 bound significantly under these conditions.

Using an affinity resin containing a covalently attached synthetic peptide corresponding to the 56 amino acid minimal activation domain of NTF-1, we showed that this region is sufficient to interact with dTAFII150 and that the activator interface of dTAFII150 is distinct from the C-terminal region with interacts with dTBP and dTAFII250. Hence, the requirement for TAFs during NTF-1 activation is at least in part mediated by NTF-1:dTAFII150 interactions.

TAF Sequence Data

Nucleotide and amino acid sequences of:

dTAFII30α.(SEQ ID NO:21, 22)
dTAFII30β.(SEQ ID NO:23, 24)
dTAFII40 (SEQ ID NO:8, 9)
dTAFII60 (SEQ ID NO:6, 7)
dTAFII80 (SEQ ID NO:4, 5)
dTAFII110 (SEQ ID NO:1, 2)
dTAF150 (SEQ ID NO:19, 20)
dTAFII250 (SEQ ID NO:3, 14)

hTAFII30α.(SEQ ID NO:28)
hTAFII30β.(SEQ ID NO:27)
hTAFII40 (SEQ ID NO:25, 26)
hTAFII70 (SEQ ID NO:12, 13)
hTAFII100 (SEQ ID NO:17, 18)
hTAFII130 (SEQ ID NO:15, 16)
hTAFII250 (SEQ ID NO:10, 11)

hTAF148 (SEQ ID NO:29, 30)
hTAFII10 (SEQ ID NO:31, 32)

were obtained as described above. Additional methods relating to PolI TAFs may be found in Comai et al. (1992) Cell 68, 965-976.

It is evident from the above results that one can use the methods and compositions disclosed herein for making and identifying diagnostic probes and therapeutic drugs. It will also be clear to one skilled in the art from a reading of this disclosure that advantage can be taken to effect alterations of gene expression: both genes encoding TAF and genes amenable to TAF-mediated transcriptional modulation. Such alterations can be effected for example, using a small molecule drug identified with disclosed TAF-based screening assays.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Tjian, Robert
Comai, Lucio
Dynlacht, Brian D
Hoey, Timothy
Ruppert, Siegfried
Tanese, Naoko
Wang, Edith
Weinzierl, Robert O.J.

(ii) TITLE OF INVENTION: TATA-Binding Protein Associated Factors,
nucleic acids encoding TAFs, and Methods of Use.

(iii) NUMBER OF SEQUENCES: 32

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
(B) STREET: 4 Embarcadero Center, 34th Floor
(C) CITY: San Francisco
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US94/
(B) FILING DATE: 28-JAN-1994
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Osman, Richard A
(B) REGISTRATION NUMBER: 36,627
(C) REFERENCE/DOCKET NUMBER: FP57650-2RA0

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-494-8700
(B) TELEFAX: 415-494-8771

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4615 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 538..3300

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAACTCGTCC GTACCTCGGC GGTCCGTAAA CAATATTTAC TCGGTTTTCG GCTAAATCGC
60

CAGAGAAACG CAACGGGAAA TCGTTTAAAA TGCGCCCCAG TGCACCGAGT TTGAACGCAA
120

AATGAATTGA ATGCTCAACA ATCAGTCCGT GCGAGCACGC GCGAGTGTGT GTGTGCGCAG
180

GAAAACCCGC CGATCGGGAA AAGTGTAGAA AGGCTTAGCG GCGCAAACAA AAGGCAGCGA
240

ATTAGCGAGA TAACACACAC GCGACAACGA CTGCAACGGA TGCGCCAGGA GAAAGGCCGA
300

CGACAGTGAC GGCAAAGGCG AGTGCGAGTG AGCCAGCGCA GCACCAATTC AGCGGAGCAC
360

CCGCTTTTTT GGCCAAGTTC GCTTCTGGAG CGCACAGCAT GCAACAACTC CGCCAACACC
420

AACACAGGAT GTGCGCAACT AGTTGATCGG AACAGGATCG CTCGCCCACA CCAACACACA
480

GAAGTCAGTG GAATAGGAGA AACACACTCG CCAATAACAT AAACACCACA CAGCACG
537

ATG AAC ACC AGC CAG ACA GCT GCC GGC AAT CGC ATC ACC TTC ACC AGC
585

```
Met Asn Thr Ser Gln Thr Ala Ala Gly Asn Arg Ile Thr Phe Thr
Ser
  1               5                  10
15

CAG CCG CTG CCC AAT GGC ACC ATC AGC ATA GCC GGC AAT CCC GGC
GCG
   633
Gln Pro Leu Pro Asn Gly Thr Ile Ser Ile Ala Gly Asn Pro Gly
Ala
            20                  25                  30

GTC ATC TCC ACG GCC CAG CTA CCG AAT ACC ACC ACC ATC AAG ACG
ATC
   681
Val Ile Ser Thr Ala Gln Leu Pro Asn Thr Thr Thr Ile Lys Thr
Ile
         35                  40                  45

CAG GCG GGG ATC GGT GGT CAG CAT CAG GGA CTT CAG CAG GTG CAT
CAT
   729
Gln Ala Gly Ile Gly Gly Gln His Gln Gly Leu Gln Gln Val His
His
     50                  55                  60

GTC CAA CAG CAG CAG CAG TCG CAA CAG CAA CAA CAG CAG CAA CAG
CAG
   777
Val Gln Gln Gln Gln Gln Ser Gln Gln Gln Gln Gln Gln Gln Gln
Gln
 65                  70                  75
   80

ACG CAA TCC GCC GGT CAA CCG CTG CTC AAT TCA ATG CTG CCG GCT
GGC
   825
Thr Gln Ser Ala Gly Gln Pro Leu Leu Asn Ser Met Leu Pro Ala
Gly
                85                  90
95

GTG GTG GTG GGC ATG CGC CAA CAG GCG CCG TCA CAG CAG CAG CAG
AAG
   873
Val Val Val Gly Met Arg Gln Gln Ala Pro Ser Gln Gln Gln Gln
Lys
            100                 105                 110
```

```
AAT GTG CCC ACC AAC CCG CTC AGT CGC GTG GTG ATC AAC TCC CAC
ATG
   921
Asn Val Pro Thr Asn Pro Leu Ser Arg Val Val Ile Asn Ser His
Met
       115                 120                 125


GCG GGC GTG AGA CCG CAG AGT CCA TCG ATA ACT TTA AGC ACA CTT
AAT
   969
Ala Gly Val Arg Pro Gln Ser Pro Ser Ile Thr Leu Ser Thr Leu
Asn
    130                 135                 140


ACG GGT CAG ACC CCG GCA TTG CTG GTC AAG ACG GAT AAC GGA TTC
CAG
  1017
Thr Gly Gln Thr Pro Ala Leu Leu Val Lys Thr Asp Asn Gly Phe
Gln
145                 150                 155
  160

CTG TTG CGC GTG GGC ACG ACG ACG GGT CCG CCG ACG GTG ACA CAG
ACT
  1065
Leu Leu Arg Val Gly Thr Thr Thr Gly Pro Pro Thr Val Thr Gln
Thr
                165                 170                 175

ATA ACC AAC ACC AGC AAT AAC AGC AAC ACG ACA AGC ACC ACA AAC
CAT
  1113
Ile Thr Asn Thr Ser Asn Asn Ser Asn Thr Thr Ser Thr Thr Asn
His
            180                 185                 190

CCC ACA ACC ACA CAG ATC CGT CTG CAA ACT GTG CCG GCT GCA GCT
TCT
  1161
Pro Thr Thr Thr Gln Ile Arg Leu Gln Thr Val Pro Ala Ala Ala
Ser
        195                 200                 205

ATG ACC AAC ACG ACC GCC ACC AGC AAC ATC ATT GTC AAT TCG GTG
GCA
  1209
Met Thr Asn Thr Thr Ala Thr Ser Asn Ile Ile Val Asn Ser Val
Ala
```

```
    210                 215                 220

AGC AGT GGA TAT GCA AAC TCT TCG CAG CCG CCG CAT CTG ACG CAA
CTA
  1257
Ser Ser Gly Tyr Ala Asn Ser Ser Gln Pro Pro His Leu Thr Gln
Leu
225                 230                 235
  240

AAT GCG CAG GCG CCA CAA CTG CCG CAG ATT ACG CAG ATT CAA ACA
ATA
  1305
Asn Ala Gln Ala Pro Gln Leu Pro Gln Ile Thr Gln Ile Gln Thr
Ile
                245                 250                 255

CCG GCC CAG CAG TCT CAG CAG CAG CAG GTG AAC AAT GTA AGC TCC
GCG
  1353
Pro Ala Gln Gln Ser Gln Gln Gln Gln Val Asn Asn Val Ser Ser
Ala
            260                 265                 270

GGA GGA ACG GCA ACG GCG GTC AGC AGT ACG ACG GCA GCG ACG ACG
ACG
  1401
Gly Gly Thr Ala Thr Ala Val Ser Ser Thr Thr Ala Ala Thr Thr
Thr
        275                 280                 285

CAG CAG GGC AAT ACC AAA GAA AAG TGT CGC AAG TTT CTA GCC AAT
TTA
  1449
Gln Gln Gly Asn Thr Lys Glu Lys Cys Arg Lys Phe Leu Ala Asn
Leu
    290                 295                 300

ATC GAA TTG TCG ACA CGG GAA CCG AAG CCG GTG GAG AAG AAC GTG
CGC
  1497
Ile Glu Leu Ser Thr Arg Glu Pro Lys Pro Val Glu Lys Asn Val
Arg
305                 310                 315
  320

ACC CTC ATC CAG GAG CTG GTC AAT GCG AAT GTC GAG CCG GAG GAG
TTT
  1545
```

Thr Leu Ile Gln Glu Leu Val Asn Ala Asn Val Glu Pro Glu Glu Phe
325 330 335

TGT GAC CGC CTG GAG CGC TTG CTC AAC GCC AGC CCG CAG CCG TGT TTG
1593
Cys Asp Arg Leu Glu Arg Leu Leu Asn Ala Ser Pro Gln Pro Cys Leu
340 345 350

ATT GGA TTC CTT AAG AAG AGT TTG CCT CTG CTA CGA CAA GCC CTC TAC
1641
Ile Gly Phe Leu Lys Lys Ser Leu Pro Leu Leu Arg Gln Ala Leu Tyr
355 360 365

ACA AAG GAG CTG GTC ATC GAA GGC ATT AAA CCT CCG CCG CAG CAC GTT
1689
Thr Lys Glu Leu Val Ile Glu Gly Ile Lys Pro Pro Pro Gln His Val
370 375 380

CTC GGC CTG GCC GGA CTC TCT CAA CAG TTG CCT AAA ATC CAA GCG CAA
1737
Leu Gly Leu Ala Gly Leu Ser Gln Gln Leu Pro Lys Ile Gln Ala Gln
385 390 395
400

ATC CGT CCG ATC GGT CCT AGC CAG ACA ACG ACC ATT GGA CAG ACG CAG
1785
Ile Arg Pro Ile Gly Pro Ser Gln Thr Thr Thr Ile Gly Gln Thr Gln
405 410 415

GTG CGT ATG ATA ACG CCG AAT GCC TTG GGC ACG CCG CGA CCC ACC ATT
1833
Val Arg Met Ile Thr Pro Asn Ala Leu Gly Thr Pro Arg Pro Thr Ile
420 425 430

```
GGC CAC ACC ACG ATA TCG AAG CAG CCA CCG AAT ATT CGG TTG CCT
ACG
  1881
Gly His Thr Thr Ile Ser Lys Gln Pro Pro Asn Ile Arg Leu Pro
Thr
        435                 440                 445

GCC CCG CGT CTC GTC AAC ACT GGA GGA ATT CGC ACC CAG ATA CCC
TCG
  1929
Ala Pro Arg Leu Val Asn Thr Gly Gly Ile Arg Thr Gln Ile Pro
Ser
    450                 455                 460

TTG CAG GTG CCT GGT CAG GCG AAC ATT GTG CAA ATA CGT GGA CCG
CAG
  1977
Leu Gln Val Pro Gly Gln Ala Asn Ile Val Gln Ile Arg Gly Pro
Gln
465                 470                 475
  480

CAT GCT CAG CTG CAG CGT ACT GGA TCG GTC CAG ATC CGG GCC ACC
ACT
  2025
His Ala Gln Leu Gln Arg Thr Gly Ser Val Gln Ile Arg Ala Thr
Thr
                485                 490                 495

CGT CCG CCA AAC AGT GTG CCC ACC GCG AAC AAA CTC ACT GCC GTC
AAG
  2073
Arg Pro Pro Asn Ser Val Pro Thr Ala Asn Lys Leu Thr Ala Val
Lys
            500                 505                 510

GTG GGA CAG ACG CAA ATC AAA GCG ATT ACG CCC AGC CTG CAT CCA
CCC
  2121
Val Gly Gln Thr Gln Ile Lys Ala Ile Thr Pro Ser Leu His Pro
Pro
        515                 520                 525

TCG CTG GCG GCA ATC TCA GGT GGA CCA CCG CCG ACA CCC ACG CTG
TCT
  2169
Ser Leu Ala Ala Ile Ser Gly Gly Pro Pro Pro Thr Pro Thr Leu
Ser
```

```
    530                 535                 540

GTT TTG TCT ACG TTG AAC TCC GCC TCG ACC ACA ACG CTG CCC ATA
CCA
  2217
Val Leu Ser Thr Leu Asn Ser Ala Ser Thr Thr Thr Leu Pro Ile
Pro
545                 550                 555
  560

TCG TTA CCC ACG GTC CAC CTT CCC CCC GAA GCT CTT CGA GCC CGT
GAG
  2265
Ser Leu Pro Thr Val His Leu Pro Pro Glu Ala Leu Arg Ala Arg
Glu
                565                 570                 575

CAG ATG CAA AAT TCG CTG AAC CAC AAC AGC AAT CAC TTC GAT GCA
AAA
  2313
Gln Met Gln Asn Ser Leu Asn His Asn Ser Asn His Phe Asp Ala
Lys
            580                 585                 590

CTG GTG GAG ATC AAG GCG CCG TCG CTG CAT CCG CCG CAC ATG GAG
CGG
  2361
Leu Val Glu Ile Lys Ala Pro Ser Leu His Pro Pro His Met Glu
Arg
        595                 600                 605

ATC AAC GCA TCT CTC ACA CCG ATT GGA GCC AAG ACG ATG GCA AGG
CCG
  2409
Ile Asn Ala Ser Leu Thr Pro Ile Gly Ala Lys Thr Met Ala Arg
Pro
    610                 615                 620

CCG CCT GCG ATC AAC AAG GCG ATA GGG AAA AAG AAA CGC GAC GCC
ATG
  2457
Pro Pro Ala Ile Asn Lys Ala Ile Gly Lys Lys Lys Arg Asp Ala
Met
625                 630                 635
  640

GAA ATG GAC GCC AAA TTG AAC ACA TCG AGC GGA GGA GCG GCG TCC
GCT
  2505
```

```
Glu Met Asp Ala Lys Leu Asn Thr Ser Ser Gly Gly Ala Ala Ser
Ala
                645                 650                 655

GCG AAC TCG TTT TTC CAG CAG AGC TCC ATG TCC TCG ATG TAC GGT
GAC
  2553
Ala Asn Ser Phe Phe Gln Gln Ser Ser Met Ser Ser Met Tyr Gly
Asp
            660                 665                 670

GAT GAT ATC AAC GAT GTT GCC GCC ATG GGA GGT GTT AAC TTG GCG
GAG
  2601
Asp Asp Ile Asn Asp Val Ala Ala Met Gly Gly Val Asn Leu Ala
Glu
        675                 680                 685

GAG TCG CAG CGA ATT CTC GGC TGT ACC GAA AAC ATC GGC ACG CAG
ATT
  2649
Glu Ser Gln Arg Ile Leu Gly Cys Thr Glu Asn Ile Gly Thr Gln
Ile
    690                 695                 700

CGA TCC TGC AAA GAT GAG GTT TTT CTT AAT CTC CCC TCG CTG CAA
GCT
  2697
Arg Ser Cys Lys Asp Glu Val Phe Leu Asn Leu Pro Ser Leu Gln
Ala
705                 710                 715
  720

AGA ATA CGG GCA ATT ACT TCG GAG GCG GGA CTG GAT GAG CCG TCG
CAG
  2745
Arg Ile Arg Ala Ile Thr Ser Glu Ala Gly Leu Asp Glu Pro Ser
Gln
                725                 730                 735

GAT GTG GCC GTT CTG ATA TCG CAC GCC TGT CAG GAG CGC CTG AAG
AAC
  2793
Asp Val Ala Val Leu Ile Ser His Ala Cys Gln Glu Arg Leu Lys
Asn
            740                 745                 750
```

```
ATC GTT GAG AAG TTG GCT GTG ATA GCG GAG CAC CGC ATT GAT GTC
ATC
  2841
Ile Val Glu Lys Leu Ala Val Ile Ala Glu His Arg Ile Asp Val
Ile
        755                 760                 765


AAG TTG GAT CCA CGC TAT GAG CCC GCC AAG GAT GTG CGC GGT CAG
ATC
  2889
Lys Leu Asp Pro Arg Tyr Glu Pro Ala Lys Asp Val Arg Gly Gln
Ile
    770                 775                 780


AAG TTT CTC GAG GAG CTG GAC AAG GCC GAG CAG AAG CGA CAC GAG
GAA
  2937
Lys Phe Leu Glu Glu Leu Asp Lys Ala Glu Gln Lys Arg His Glu
Glu
785                 790                 795
  800

CTG GAG CGT GAG ATG CTG CTG CGG GCA GCC AAG TCA CGG TCG AGG
GTG
  2985
Leu Glu Arg Glu Met Leu Leu Arg Ala Ala Lys Ser Arg Ser Arg
Val
                805                 810                 815

GAA GAT CCC GAG CAG GCC AAG ATG AAG GCG AGG GCC AAG GAG ATG
CAA
  3033
Glu Asp Pro Glu Gln Ala Lys Met Lys Ala Arg Ala Lys Glu Met
Gln
            820                 825                 830

CGC GCC GAA ATG GAG GAG TTG CGT CAA CGA GAT GCC AAT CTG ACG
GCG
  3081
Arg Ala Glu Met Glu Glu Leu Arg Gln Arg Asp Ala Asn Leu Thr
Ala
        835                 840                 845

CTG CAG GCG ATT GGA CCT CGG AAA AAG CTG AAG CTG GAC GGC GAA
ACA
  3129
Leu Gln Ala Ile Gly Pro Arg Lys Lys Leu Lys Leu Asp Gly Glu
Thr
```

```
     850                 855                 860

GTC AGT TCG GGA GCG GGT TCA AGT GGC GGC GGA GTG CTA AGC AGC
TCG
  3177
Val Ser Ser Gly Ala Gly Ser Ser Gly Gly Gly Val Leu Ser Ser
Ser
865                 870                 875
  880

GGA TCT GCG CCG ACG ACG TTA CGG CCT CGC ATA AAA CGT GTG AAC
CTG
  3225
Gly Ser Ala Pro Thr Thr Leu Arg Pro Arg Ile Lys Arg Val Asn
Leu
               885                 890                 895


CGC GAC ATG CTC TTC TAC ATG GAG CAA GAG CGG GAG TTC TGT CGC
AGT
  3273
Arg Asp Met Leu Phe Tyr Met Glu Gln Glu Arg Glu Phe Cys Arg
Ser
            900                 905                 910


TCC ATG CTG TTC AAG ACA TAC CTC AAG TGATCGCTGC TGTTGCCCAT

  3320
Ser Met Leu Phe Lys Thr Tyr Leu Lys

        915                 920


CAATCGCACC GTCTTCTCCT CGCCGATCCT CCTACTCCGT GGACTGTCGT
GTTGTTGTTT
  3380

TATACAGCTT TACGATTTCA TCCACTTGCA ATATATTTTA GCCTCAACTT
TAAATGCGTC
  3440

GCGTGTCCCC TGTTGTTGTT TCTTTTTAGT TAGGCGGCTC TATTTAATTT
CTATTTTTAC
  3500

ATTTATTTAC ATAAATCCTA AATTCTAATC GTATTTGATT TTAAGCCTAA
TTTAAAGCTC
  3560

GTTTATTTTT CCAATAAATT CTCTGTAAAA CTTAAACCAA ACCAATCCAA
AAACAAAACA
  3620
```

AAACCAGAGT AAACGAAGAG AATAAAATAA TAGAGAGGAA AGTAAAAGAA
GGTAAAAGAG
3680

AGCGCGCAGT CAGCGGTCGT TTGATTTGTA ATTTGTAACA TAATAATGTT
TGCATCAACT
3740

GCATTGACGG CCTTATCTAA ACGATATAAA CATAATTATT AATATTTAAT
TATTTAGCTT
3800

AGTTTGTTAA ACGAAAACGA ACCATAATTC CTAGATTTTA AGTAAAAAGC
AAGGGCGCGT
3860

GAAGAGAAAT CGAAACCGAA TTACAGATAA AGGTTTTTAA AACCAACTAG
ATCGAAACAA
3920

GTTCAGCAAC AGCAAAACAA AAGAACACAT CAAAAAAAGA ACCGAAAAAT
ATCCATTTAA
3980

ACATCCATTG AATTAGGTTT AGTTGTTTAA AAAAGATGTA ATTTTTAATT
ACCCATAATG
4040

TATAAACGGA AATCAATCGT TAGGCAAGAC CACAACAAAC CCAACAAATT
GTAAATACAT
4100

TCTAGGCTAC GGTTTTTCTA ATAGATAACT AGGTAAAAAC GCAAACGTAA
TTAACAAATT
4160

ATCGATGGCA AGGAGCGATG CGAGCGCAGA CAACTTGGCA CACCGAAAAA
ATATGTTTTT
4220

ATTAGTGGCG CTCGTTCATC CATTAAGAAT GGCGATTCAT TAGGCTCCAT
AGATCCATAA
4280

ATCCCCTAAT CCAATCTGAA CTACACACAA AATAGACAAA TTTTATACAA
TTAGCTCGAT
4340

AAATCTTGTA AAATAGAGTC CCGTAAAAAA TTATAACAAA TAAATTGACA
ACAATTGATG
4400

TAATTCAGTA AACCTAAGCA AAAAGTGAAA CCATTCTAAG CAAATTCTTT
GTGTGTAAAA

4460

ATTAATATGA TAAACAAAAT GCAGATGCAA CCGTAAACAG CGCATAGTTT GGTAGGCATA
4520

TAACTGAATA TATATATATT ATTATTATTA TGTTTTAACA TTAAGCAAAA AAATAAAAGA
4580

AAAAATTGAG AAAACTTCAA AAAAAAAAAA AAAAA

4615

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 921 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Thr Ser Gln Thr Ala Ala Gly Asn Arg Ile Thr Phe Thr Ser
1 5 10 15

Gln Pro Leu Pro Asn Gly Thr Ile Ser Ile Ala Gly Asn Pro Gly Ala
20 25 30

Val Ile Ser Thr Ala Gln Leu Pro Asn Thr Thr Thr Ile Lys Thr Ile
35 40 45

Gln Ala Gly Ile Gly Gly Gln His Gln Gly Leu Gln Gln Val His His
50 55 60

Val Gln Gln Gln Gln Gln Ser Gln Gln Gln Gln Gln Gln Gln Gln Gln
65 70 75 80

Thr Gln Ser Ala Gly Gln Pro Leu Leu Asn Ser Met Leu Pro Ala Gly

85 90
95

Val Val Val Gly Met Arg Gln Gln Ala Pro Ser Gln Gln Gln Gln
Lys
100 105 110

Asn Val Pro Thr Asn Pro Leu Ser Arg Val Val Ile Asn Ser His
Met
115 120 125

Ala Gly Val Arg Pro Gln Ser Pro Ser Ile Thr Leu Ser Thr Leu
Asn
130 135 140

Thr Gly Gln Thr Pro Ala Leu Leu Val Lys Thr Asp Asn Gly Phe
Gln
145 150 155
160

Leu Leu Arg Val Gly Thr Thr Thr Gly Pro Pro Thr Val Thr Gln
Thr
165 170 175

Ile Thr Asn Thr Ser Asn Asn Ser Asn Thr Thr Ser Thr Thr Asn
His
180 185 190

Pro Thr Thr Thr Gln Ile Arg Leu Gln Thr Val Pro Ala Ala Ala
Ser
195 200 205

Met Thr Asn Thr Thr Ala Thr Ser Asn Ile Ile Val Asn Ser Val
Ala
210 215 220

Ser Ser Gly Tyr Ala Asn Ser Ser Gln Pro Pro His Leu Thr Gln
Leu
225 230 235
240

Asn Ala Gln Ala Pro Gln Leu Pro Gln Ile Thr Gln Ile Gln Thr
Ile
245 250 255

Pro Ala Gln Gln Ser Gln Gln Gln Val Asn Asn Val Ser Ser Ala
260 265 270

Gly Gly Thr Ala Thr Ala Val Ser Ser Thr Thr Ala Ala Thr Thr Thr
275 280 285

Gln Gln Gly Asn Thr Lys Glu Lys Cys Arg Lys Phe Leu Ala Asn Leu
290 295 300

Ile Glu Leu Ser Thr Arg Glu Pro Lys Pro Val Glu Lys Asn Val Arg
305 310 315 320

Thr Leu Ile Gln Glu Leu Val Asn Ala Asn Val Glu Pro Glu Glu Phe
325 330 335

Cys Asp Arg Leu Glu Arg Leu Leu Asn Ala Ser Pro Gln Pro Cys Leu
340 345 350

Ile Gly Phe Leu Lys Lys Ser Leu Pro Leu Leu Arg Gln Ala Leu Tyr
355 360 365

Thr Lys Glu Leu Val Ile Glu Gly Ile Lys Pro Pro Pro Gln His Val
370 375 380

Leu Gly Leu Ala Gly Leu Ser Gln Gln Leu Pro Lys Ile Gln Ala Gln
385 390 395 400

Ile Arg Pro Ile Gly Pro Ser Gln Thr Thr Thr Ile Gly Gln Thr Gln
405 410 415

Val Arg Met Ile Thr Pro Asn Ala Leu Gly Thr Pro Arg Pro Thr Ile
420 425 430

Gly His Thr Thr Ile Ser Lys Gln Pro Pro Asn Ile Arg Leu Pro Thr
435 440 445

Ala Pro Arg Leu Val Asn Thr Gly Gly Ile Arg Thr Gln Ile Pro Ser
450 455 460

Leu Gln Val Pro Gly Gln Ala Asn Ile Val Gln Ile Arg Gly Pro Gln
465 470 475 480

His Ala Gln Leu Gln Arg Thr Gly Ser Val Gln Ile Arg Ala Thr Thr
485 490 495

Arg Pro Pro Asn Ser Val Pro Thr Ala Asn Lys Leu Thr Ala Val Lys
500 505 510

Val Gly Gln Thr Gln Ile Lys Ala Ile Thr Pro Ser Leu His Pro Pro
515 520 525

Ser Leu Ala Ala Ile Ser Gly Gly Pro Pro Pro Thr Pro Thr Leu Ser
530 535 540

Val Leu Ser Thr Leu Asn Ser Ala Ser Thr Thr Thr Leu Pro Ile Pro
545 550 555 560

Ser Leu Pro Thr Val His Leu Pro Pro Glu Ala Leu Arg Ala Arg Glu
565 570 575

Gln Met Gln Asn Ser Leu Asn His Asn Ser Asn His Phe Asp Ala Lys
580 585 590

Leu Val Glu Ile Lys Ala Pro Ser Leu His Pro Pro His Met Glu Arg
595 600 605

Ile Asn Ala Ser Leu Thr Pro Ile Gly Ala Lys Thr Met Ala Arg Pro
610 615 620

Pro Pro Ala Ile Asn Lys Ala Ile Gly Lys Lys Lys Arg Asp Ala Met
625 630 635 640

Glu Met Asp Ala Lys Leu Asn Thr Ser Ser Gly Gly Ala Ala Ser Ala
645 650 655

Ala Asn Ser Phe Phe Gln Gln Ser Ser Met Ser Ser Met Tyr Gly Asp
660 665 670

Asp Asp Ile Asn Asp Val Ala Ala Met Gly Gly Val Asn Leu Ala Glu
675 680 685

Glu Ser Gln Arg Ile Leu Gly Cys Thr Glu Asn Ile Gly Thr Gln Ile
690 695 700

Arg Ser Cys Lys Asp Glu Val Phe Leu Asn Leu Pro Ser Leu Gln Ala
705 710 715 720

Arg Ile Arg Ala Ile Thr Ser Glu Ala Gly Leu Asp Glu Pro Ser Gln
725 730 735

Asp Val Ala Val Leu Ile Ser His Ala Cys Gln Glu Arg Leu Lys Asn
740 745 750

Ile Val Glu Lys Leu Ala Val Ile Ala Glu His Arg Ile Asp Val Ile
755 760 765

Lys Leu Asp Pro Arg Tyr Glu Pro Ala Lys Asp Val Arg Gly Gln Ile
770 775 780

Lys Phe Leu Glu Glu Leu Asp Lys Ala Glu Gln Lys Arg His Glu Glu
785 790 795 800

Leu Glu Arg Glu Met Leu Leu Arg Ala Ala Lys Ser Arg Ser Arg Val
805 810 815

Glu Asp Pro Glu Gln Ala Lys Met Lys Ala Arg Ala Lys Glu Met Gln
820 825 830

Arg Ala Glu Met Glu Glu Leu Arg Gln Arg Asp Ala Asn Leu Thr Ala
835 840 845

Leu Gln Ala Ile Gly Pro Arg Lys Lys Leu Lys Leu Asp Gly Glu Thr
850 855 860

Val Ser Ser Gly Ala Gly Ser Ser Gly Gly Gly Val Leu Ser Ser Ser
865 870 875 880

Gly Ser Ala Pro Thr Thr Leu Arg Pro Arg Ile Lys Arg Val Asn Leu
885 890 895

Arg Asp Met Leu Phe Tyr Met Glu Gln Glu Arg Glu Phe Cys Arg Ser
900 905 910

Ser Met Leu Phe Lys Thr Tyr Leu Lys
915 920

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4164 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTCGCTGTAC GAGGTACCCG GTCCGAATTC CAAAAGGGCC AACAACTTCA CCCGTGACTT
60

TCTGCAGGTG TTTATTTACC GCCTGTTCTG GAAAAGTCGC GACAACCCGC CCCGCATTCG
120

AATGGACGAT ATAAAACAGG CTTTTCCCGC TCATTCCGAG AGCAGCATCC GCAAGCGTTT
180

AAAGCAGTGC GCTGACTTCA AGCGAACAGG CATGGACTCC AATTGGTGGG TTATAAAGCC
240

AGAGTTTCGC CTTCCATCCG AGGAGGAGAT CCGAGCCATG GTGTCACCTG AGCAGTGTTG
300

CGGTACTTCA GCATGATAGC GGCGGAACAA CGCTTAAAGG ATGCTGGGTA TGGAGAAAAG
360

TTTTTGTTCG CACCTCAGGA AGATGACGAC GAGGAGGCGC AGTGAAAGCT TGACGACGAA
420

GTAAAGGTGG CTCCTTGGAA CACGACTCGC GCATATATCC AAGCCATGCG GGGAAAGTGT
480

TTACTCCAGT TGAGTGGTCC AGCCGATCCA ACGGGATGTG GAGAGGGATT TTCATATGTT
540

CGAGTGCCAA ACAAGCCCAC GCAAACCAAG GAGGAGCAAG AGTCGCAGCC TAAACGTTCG
600

GTCACAGGAA CAGATGCAGA TTTGCGTCGT CTGCCACTCC AGCGTGCAAA AGAGCTGTTG
660

CGGCAGTTCA AGGTGCCCGA GGAGGAGATC AAAAAGCTTT CCCGCTGGGA GGTCATTGAC
720

GTGGTGCGCA CCCTGTCCAC AGAAAAGGCC AAGGCCGGTG AAGAGGGAAT GGATAAGTTT
780

TCTCGTGGCA ACCGGTTCTC CATTGCAGAG CATCAGGAGC GTTATAAGGA
AGAGTGCCAG
840

CGCATATTCG ATCTGCAAAA CAGAGTGCTG GCCAGCTCTG AGGTGCTGTC
CACAGATGAG
900

GCAGAGTCCT CGGCCTCTGA GGAATCTGAT CTCGAAGAAC TTGGCAAGAA
TCTTGAGAAC
960

ATGCTGTCAA ACAAGAAAAC CTCGACGCAA TTGTCAAGGG AACGTGAAGA
GCTGGAGCGT
1020

CAGGAGTTGC TTCGCCAGCT TGACGAAGAA CACGGCGGAC CAAGTGGTAG
TGGAGGAGCC
1080

AAGGGAGCCA AAGGAAAGGA TGATCCGGGA CAGCAAATGC TGGCAACCAA
CAACCAGGGC
1140

AGGATCCTTC GCATTACGCG TACCTTTAGA GGTAACGATG GCAAGGAATA
TACTCGCGTG
1200

GAGACTGTGC GGCGGCAACC AGTTATCGAC GCCTACATCA AGATTCGCAC
CACTAAGGAC
1260

GAGCAGTTCA TCAAGCAGTT CGCAACGCTA GATGAGCAGC AGAAGGAGGA
GATGAAGCGC
1320

GAAAAGAGAC GCATTCAGGA GCAGCTACGT CGCATCAAGC GCAACCAGGA
GCGCGAACGC
1380

CTGGCGCAGC TGGCCCAGAA CCAGAAGCTT CAGCCAGGTG GCATGCCCAC
TTCCTTGGGT
1440

GATCCTAAGA GCTCGGGCGG TCATTCGCAC AAGGAGCGGG ATAGCGGCTA
CAAGGAGGTC
1500

AGCCCTTCGC GCAAGAAGTT CAAGCTTAAG CCAGACCTAA AGCTGAAGTG
CGGCGCCTGT
1560

GGACAGGTTG GTCACATGCG CACAAACAAA GCCTGTCCCT TGTATTCTGG
CATGCAAAGC

1620

AGTCTGTCCC AGTCGAACCC ATCTCTGGCT GACGATTTTG ACGAACAGAG
CGAAAAGGAG
1680

ATGACAATGG ATGACGATGA TCTTGTGAAT GTCGATGGCA CCAAAGTAAC
GCTCAGCAGT
1740

AAGATTCTCA AGCGTCATGG TGGTGATGAT GGCAAGCGTC GCAGCGGATC
TAGCTCTGGT
1800

TTCACCTTGA AGGTTCCCCG AGATGCGATG GGCAAGAAGA AACGCAGAGT
GGGTGGCGAT
1860

CTTCATTGTG ACTATCTGCA GCGACACAAT AAAACGGCCA ATCGCAGGCG
CACGGACCCC
1920

GTTGTGGTAC TGTCCTCTAT CCTGGAGATT ATCCATAATG AGCTGCGATC
TATGCCAGAT
1980

GTATCGCCAT TCCTGTTCCC GGTAAGCGCA AAAAAGGTTC CCGACTACTA
CCGCGTGGTG
2040

ACCAAGCCCA TGGATCTGCA AACGATGAGG GAGTATATCG CCAAAGGCTA
ACACGAGTCG
2100

CGAGATGTTC CTCGAGGATC TCAAGCAGAT TGTGGACAAC TCGCTGATCT
ACAATGGACC
2160

GCAGAGTGCA TACACCTTGG CTGCCCAACG CATGTTCAGC AGTTGTTTTG
AATTGCTCGC
2220

AGAGGCGAAG ACAAACTGAT GCGCCTCGAG AAGGCAATTA ACCCGCTGCT
GGACGACGAT
2280

GACCAAGTGG CACTCTCCTT TATCTTTGAC AAGCTGCACT CGCAGATTAA
GCAATTACCA
2340

GAGAGCTGGC CTTTCCTTAA GCCTGTCAAC AAGAAACAGG TTAAGGACTA
CTACACGGTT
2400

ATCAAGCGAC CCATGGACCT CGAAACTATC GGCAAAAACA TTGAAGCTCA
TCGCTATCAC
2460

AGTCGTGCCG AGTATCTGGC TGATATCGAG TTGATCGCCA CCAACTGTGA
GCAGTACAAC
2520

GGCAGTGACA CCCGCTACAC CAAGTTCTCA AAGAAGATAC TTGAGTATGC
CCAAACCCAG
2580

TTAATTGAGT TTTCGGAGCA CTGCGGCCAG TTGGAAAATA ACATAGCTAA
GACGCAGGAG
2640

CGTGCTAGGG AAAATGCACC AGAGTTTGAT GAAGCCTGGG GCAATGATGA
TTACAACTTT
2700

GACCGTGGCA GTAGGGCCAG TTCACCCGGA GATGACTACA TCGACGTCGA
GGGTCATGGG
2760

GGGCATGCCT CCTCATCGAA CTCTATCCAT CGCAGCATGG GCGCCGAGGC
CGGTTCGTCA
2820

CATACGGCGC CGGCGGTGCG AAAACCAGCT CCTCCTGGTC CTGGTGAGGT
GAAGCGCGGA
2880

AGGGGTAGGC CCCGCAAGCA GCGCGACCCC GTGGAGGAGG TCAAATCCCA
GAATCCGGTT
2940

AAGCGTGGTC GGGGGCGTCC GAGGAAGGAC AGCCTTGCCT CAAACATGAG
TCACACGCAA
3000

GCTTACTTCC TGGATGAAGA TCTCCAATGC TCCACAGATG ACGAGGACGA
CGACGAGGAG
3060

GAGGACTTCC AGGAGGTCTC CGAAGACGAG AACAATGCGG CGAGCATTTT
AGATCAGGGC
3120

GAACGTATCA ATGCGCCTGC CGATGCCATG GATGGCATGT TTGACCCCAA
GAACATCAAG
3180

ACAGAGATTG ACCTAGAGGC TCACCAGATG GCAGAGGAGC CGATCGGCGA
GGATGACAGC

THIS PAGE IS BLANK

3240

CAGCAGGTGG CCGAAGCAAT GGTGCAGTTG AGTGGCGTGG GCGGCTACTA TGCTCAACAG
3300

CAGCAAGATG AATCCATGGA TGTGGACCCC AACTACGATC CCTCAGATTT CCTCGCCATG
3360

CACAAGCAGC GCCAGAGCCT CGGCGAGCCC AGCAGCTTGC AGGGTGCTTT CACCAACTTC
3420

CTATCGCACG AGCAGGATGA TAATGGGCCT TACAATCCCG CCGAAGCCAG CACAAGTGCC
3480

GCTTCCGGTG CAGACTTAGG AATGGACGCT TCAATGGCCA TGCAAATGGC GCCGGAAATG
3540

CCTGTCAATA CCATGAACAA CGGAATGGGC ATCGATGATG ATCTGGATAT TTCGGAGAGT
3600

GACGAGGAAG ACGATGGTTC TCGAGTGCGT ATCAAAAAGG AGGTCTTCGA CGACGGGGAT
3660

TACGCCTTGC AGCACCAGCA GATGGGACAG GCAGCATCGC AGTCGCAGAT ATACATGGGG
3720

ATTCGTCCAA CGAGCCCACG ACTCTCGACT ACCAGCAACC ACCGCAACTG GACTTCCAAC
3780

AAGTGCAGGA AATGGAGCAG TTGCAGCACC AAGTGATGCC ACCAATGCAA TCAGAGCAAC
3840

Terminate TGCAGCAGCA ACAGACGCCG CAGGAGACAA TGATTATGCC TGGACTTTTT AGTGATAGGG
3900

AATAATTGTT AGTTGTTAGA AAATAAAACG TCGATTTAAT AATAGGATTG AGCTTCGCTG
3960

TGAAACAATT TTATACACTT TTTACAATGC ATTGTTTTAA CGGATTTTGA AATACTACAA
4020

TATGTTCTCT GAAAAAATAT TTCCTTTTCA TGCCAATATG TTTTTAATTT TACACTTTAC
4080

AATTTATGAA ATCTAATTCA AAATATGTTT TTAAAATATA ATTTTCATAA CTTTAAATAA
4140

TGCCTAGAAA AAAAAAAAAA AAAA

4164

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 49..2160

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATAACAAAA TAGTACACAA GTTCCATATA TTTCAATTTT CCGCGAAA ATG AGC CTG
57
Met Ser Leu
1

GAA GTG AGC AAT ATC AAC GGG GGA AAC GGT ACT CAA TTG TCC CAC GAC
105
Glu Val Ser Asn Ile Asn Gly Gly Asn Gly Thr Gln Leu Ser His Asp
5 10 15

AAG CGT GAG CTG CTA TGC CTG CTG AAA CTC ATC AAA AAG TAC CAG CTG
153
Lys Arg Glu Leu Leu Cys Leu Leu Lys Leu Ile Lys Lys Tyr Gln Leu
20 25 30
35

```
AAG AGC ACT GAG GAG CTG CTC TGC CAA GAG GCG AAT GTG AGC AGT
GTG
   201
Lys Ser Thr Glu Glu Leu Leu Cys Gln Glu Ala Asn Val Ser Ser
Val
                40                  45
50

GAA TTG TCG GAA ATC AGC GAA AGT GAT GTT CAG CAG GTG CTG GGC
GCA
   249
Glu Leu Ser Glu Ile Ser Glu Ser Asp Val Gln Gln Val Leu Gly
Ala
            55                  60                  65

GTT TTG GGA GCT GGC GAT GCC AAC CGG GAG CGG AAA CAT GTC CAA
TCT
   297
Val Leu Gly Ala Gly Asp Ala Asn Arg Glu Arg Lys His Val Gln
Ser
        70                  75                  80

CCG GCG CAG GGT CAT AAA CAG TCC GCG GTG ACG GAG GCC AAT GCT
GCA
   345
Pro Ala Gln Gly His Lys Gln Ser Ala Val Thr Glu Ala Asn Ala
Ala
     85                  90                  95

GAG GAA CTG GCC AAG TTC ATC GAC GAC GAC AGC TTT GAT GCT CAG
CAC
   393
Glu Glu Leu Ala Lys Phe Ile Asp Asp Asp Ser Phe Asp Ala Gln
His
100                 105                 110
  115

TAT GAG CAG GCA TAC AAG GAG CTG CGC ACT TTC GTT GAG GAC TCC
CTG
   441
Tyr Glu Gln Ala Tyr Lys Glu Leu Arg Thr Phe Val Glu Asp Ser
Leu
                120                 125                 130

GAC ATA TAC AAG CAT GAG CTG TCC ATG GTT CTG TAC CCA ATT CTG
GTG
   489
Asp Ile Tyr Lys His Glu Leu Ser Met Val Leu Tyr Pro Ile Leu
Val
```

```
                135                 140                 145

CAG ATC TAC TTC AAG ATC CTC GCC AGT GGA CTA AGG GAG AAG GCC
AAA
   537
Gln Ile Tyr Phe Lys Ile Leu Ala Ser Gly Leu Arg Glu Lys Ala
Lys
        150                 155                 160

GAA TTC ATT GAG AAG TAC AAA TGC GAT CTC GAC GGC TAC TAC ATA
GAG
   585
Glu Phe Ile Glu Lys Tyr Lys Cys Asp Leu Asp Gly Tyr Tyr Ile
Glu
    165                 170                 175

GGT CTT TTC AAC CTT CTT TTG CTG TCT AAG CCC GAG GAG CTG CTG
GAG
   633
Gly Leu Phe Asn Leu Leu Leu Leu Ser Lys Pro Glu Glu Leu Leu
Glu
180                 185                 190
  195

AAT GAC CTC GTA GTA GCC ATG GAG CAG GAT AAG TTT GTC ATT CGC
ATG
   681
Asn Asp Leu Val Val Ala Met Glu Gln Asp Lys Phe Val Ile Arg
Met
                200                 205                 210

TCC AGG GAC TCG CAC TCT CTG TTC AAG CGA CAC ATT CAG GAT CGC
CGG
   729
Ser Arg Asp Ser His Ser Leu Phe Lys Arg His Ile Gln Asp Arg
Arg
            215                 220                 225

CAG GAA GTG GTG GCA GAT ATT GTT TCC AAG TAC TTG CAT TTC GAC
ACA
   777
Gln Glu Val Val Ala Asp Ile Val Ser Lys Tyr Leu His Phe Asp
Thr
        230                 235                 240

TAC GAG GGC ATG GCG CGC AAC AAG CTG CAG TGC GTC GCC ACC GCG
GGC
   825
```

```
Tyr Glu Gly Met Ala Arg Asn Lys Leu Gln Cys Val Ala Thr Ala
Gly
    245                 250                 255


TCG CAC CTC GGA GAG GCC AAG CGA CAG GAC AAC AAA ATG CGG GTG
TAC
   873
Ser His Leu Gly Glu Ala Lys Arg Gln Asp Asn Lys Met Arg Val
Tyr
260                 265                 270
  275

TAC GGA CTG CTC AAG GAG GTG GAC TTT CAG ACT CTG ACC ACT CCA
GCG
   921
Tyr Gly Leu Leu Lys Glu Val Asp Phe Gln Thr Leu Thr Thr Pro
Ala
                280                 285                 290


CCG GCA CCA GAG GAG GAG GAC GAT GAT CCG GAT GCC CCG GAT CGT
CCG
   969
Pro Ala Pro Glu Glu Glu Asp Asp Asp Pro Asp Ala Pro Asp Arg
Pro
            295                 300                 305


AAA AAG AAA AAG CCA AAA AAG GAT CCC CTG CTG TCG AAA AAG TCC
AAG
  1017
Lys Lys Lys Lys Pro Lys Lys Asp Pro Leu Leu Ser Lys Lys Ser
Lys
        310                 315                 320


TCG GAT CCG AAT GCT CCA TCC ATC GAC AGA ATT CCC CTG CCG GAA
CTG
  1065
Ser Asp Pro Asn Ala Pro Ser Ile Asp Arg Ile Pro Leu Pro Glu
Leu
    325                 330                 335


AAG GAT TCG GAC AAG TTG CTA AAG CTT AAG GCT CTC AGG GAA GCC
AGC
  1113
Lys Asp Ser Asp Lys Leu Leu Lys Leu Lys Ala Leu Arg Glu Ala
Ser
340                 345                 350
  355
```

```
AAG CGT TTA GCC CTC AGC AAG GAT CAA CTG CCC TCT GCC GTC TTC
TAC
  1161
Lys Arg Leu Ala Leu Ser Lys Asp Gln Leu Pro Ser Ala Val Phe
Tyr
                360                 365                 370


ACG GTG CTT AAT TCC CAT CAG GGC GTA ACC TGT GCC GAG ATT TCA
GAC
  1209
Thr Val Leu Asn Ser His Gln Gly Val Thr Cys Ala Glu Ile Ser
Asp
            375                 380                 385


GAT TCC ACG ATG TTG GCC TGT GGA TTT GGC GAT TCT AGC GTG AGG
ATT
  1257
Asp Ser Thr Met Leu Ala Cys Gly Phe Gly Asp Ser Ser Val Arg
Ile
        390                 395                 400


TGG TCA TTG ACG CCC GCG AAG CTG CGT ACG CTG AAG GAT GCA GAT
TCC
  1305
Trp Ser Leu Thr Pro Ala Lys Leu Arg Thr Leu Lys Asp Ala Asp
Ser
    405                 410                 415


CTT CGC GAA CTG GAC AAG GAA TCG GCG GAT ATC AAT GTG CGT ATG
CTG
  1353
Leu Arg Glu Leu Asp Lys Glu Ser Ala Asp Ile Asn Val Arg Met
Leu
420                 425                 430
  435

GAT GAC CGA AGT GGT GAG GTA ACC AGG AGC TTA ATG GGT CAC ACC
GGA
  1401
Asp Asp Arg Ser Gly Glu Val Thr Arg Ser Leu Met Gly His Thr
Gly
                440                 445                 450


CCC GTA TAC CGC TGT GCC TTT GCC CCC GAG ATG AAC CTG TTG CTC
TCA
  1449
Pro Val Tyr Arg Cys Ala Phe Ala Pro Glu Met Asn Leu Leu Leu
Ser
```

```
                455                 460                 465

TGT TCC GAG GAC AGC ACC ATA AGG CTG TGG TCT CTG CTC ACC TGG
TCC
  1497
Cys Ser Glu Asp Ser Thr Ile Arg Leu Trp Ser Leu Leu Thr Trp
Ser
        470                 475                 480

TGC GTA GTC ACC TAC CGC GGG CAC GTT TAC CCG GTG TGG GAT GTT
CGC
  1545
Cys Val Val Thr Tyr Arg Gly His Val Tyr Pro Val Trp Asp Val
Arg
    485                 490                 495

TTT GCG CCG CAT GGC TAC TAT TTT GTT TCT TGT TCG TAC GAC AAA
ACT
  1593
Phe Ala Pro His Gly Tyr Tyr Phe Val Ser Cys Ser Tyr Asp Lys
Thr
500                 505                 510
  515

GCT CGT CTG TGG GCC ACG GAT TCC AAT CAA GCG TTG CGC GTA TTC
GTG
  1641
Ala Arg Leu Trp Ala Thr Asp Ser Asn Gln Ala Leu Arg Val Phe
Val
                520                 525                 530

GGT CAC TTG TCG GAC GTG GAT TGT GTA CAA TTT CAT CCC AAT TCC
AAT
  1689
Gly His Leu Ser Asp Val Asp Cys Val Gln Phe His Pro Asn Ser
Asn
            535                 540                 545

TAT GTG GCC ACC GGT TCT AGC GAT CGC ACG GTA CGC CTG TGG GAC
AAC
  1737
Tyr Val Ala Thr Gly Ser Ser Asp Arg Thr Val Arg Leu Trp Asp
Asn
        550                 555                 560

ATG ACC GGT CAG TCG GTA CGC CTG ATG ACG GGC CAC AAG GGA TCG
GTG
  1785
```

```
Met Thr Gly Gln Ser Val Arg Leu Met Thr Gly His Lys Gly Ser
Val
    565                 570                 575

AGT TCT CTG GCC TTC TCC GCC TGC GGC CGG TAT CTG GCC TCG GGT
TCA
  1833
Ser Ser Leu Ala Phe Ser Ala Cys Gly Arg Tyr Leu Ala Ser Gly
Ser
580                 585                 590
  595

GTA GAT CAC AAT ATC ATC ATC TGG GAT CTG TCG AAC GGA TCC CTG
GTC
  1881
Val Asp His Asn Ile Ile Ile Trp Asp Leu Ser Asn Gly Ser Leu
Val
                600                 605                 610

ACC ACC CTG TTG AGG CAC ACT AGC ACT GTG ACC ACG ATC ACC TTT
AGT
  1929
Thr Thr Leu Leu Arg His Thr Ser Thr Val Thr Thr Ile Thr Phe
Ser
            615                 620                 625

CGC GAT GGA ACA GTC CTG GCT GCA GCC GGC TTG GAT AAC AAT CTA
ACT
  1977
Arg Asp Gly Thr Val Leu Ala Ala Ala Gly Leu Asp Asn Asn Leu
Thr
        630                 635                 640

CTG TGG GAC TTT CAC AAG GTT ACC GAA GAC TAT ATC AGC AAT CAC
ATC
  2025
Leu Trp Asp Phe His Lys Val Thr Glu Asp Tyr Ile Ser Asn His
Ile
    645                 650                 655

ACT GTG TCG CAC CAT CAG GAT GAG AAC GAC GAG GAC GTC TAC CTC
ATG
  2073
Thr Val Ser His His Gln Asp Glu Asn Asp Glu Asp Val Tyr Leu
Met
660                 665                 670
  675
```

CGT ACT TTC CCC AGC AAG AAC TCG CCA TTT GTC AGC CTG CAC TTT ACG
2121
Arg Thr Phe Pro Ser Lys Asn Ser Pro Phe Val Ser Leu His Phe Thr
680 685 690

CGC CGA AAT CTC CTG ATG TGC GTG GGT CTA TTC AAG AGT TAGGAGCACA
2170
Arg Arg Asn Leu Leu Met Cys Val Gly Leu Phe Lys Ser
695 700

GATAAGCTTA TTTGGTATAC GTAATGTAGT GTTAAGGAAT GCTCGGAATG TTTAGGATTA
2230

ATGTTTTGTA TTTCGTTTGT GACCCATCCC CCCTGAAATG TCGATTAGTT GTTTAAGCAT
2290

AAAAGTGTAA AGTGCATATA TGCGCAAGTT ATCAATAAAT TTTAATTAAT ATAAAAGTCA
2350

AAAAAAAAA

2359

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 704 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ser Leu Glu Val Ser Asn Ile Asn Gly Gly Asn Gly Thr Gln Leu
1 5 10 15

Ser His Asp Lys Arg Glu Leu Leu Cys Leu Leu Lys Leu Ile Lys Lys
20 25 30

Tyr Gln Leu Lys Ser Thr Glu Glu Leu Leu Cys Gln Glu Ala Asn Val
35 40 45

Ser Ser Val Glu Leu Ser Glu Ile Ser Glu Ser Asp Val Gln Gln Val
50 55 60

Leu Gly Ala Val Leu Gly Ala Gly Asp Ala Asn Arg Glu Arg Lys His
65 70 75 80

Val Gln Ser Pro Ala Gln Gly His Lys Gln Ser Ala Val Thr Glu Ala
85 90 95

Asn Ala Ala Glu Glu Leu Ala Lys Phe Ile Asp Asp Asp Ser Phe Asp
100 105 110

Ala Gln His Tyr Glu Gln Ala Tyr Lys Glu Leu Arg Thr Phe Val Glu
115 120 125

Asp Ser Leu Asp Ile Tyr Lys His Glu Leu Ser Met Val Leu Tyr Pro
130 135 140

Ile Leu Val Gln Ile Tyr Phe Lys Ile Leu Ala Ser Gly Leu Arg Glu
145 150 155 160

Lys Ala Lys Glu Phe Ile Glu Lys Tyr Lys Cys Asp Leu Asp Gly Tyr
165 170 175

Tyr Ile Glu Gly Leu Phe Asn Leu Leu Leu Leu Ser Lys Pro Glu Glu
180 185 190

Leu Leu Glu Asn Asp Leu Val Val Ala Met Glu Gln Asp Lys Phe Val
195 200 205

Ile Arg Met Ser Arg Asp Ser His Ser Leu Phe Lys Arg His Ile Gln
210 215 220

Asp Arg Arg Gln Glu Val Val Ala Asp Ile Val Ser Lys Tyr Leu His
225 230 235 240

Phe Asp Thr Tyr Glu Gly Met Ala Arg Asn Lys Leu Gln Cys Val Ala
245 250 255

Thr Ala Gly Ser His Leu Gly Glu Ala Lys Arg Gln Asp Asn Lys Met
260 265 270

Arg Val Tyr Tyr Gly Leu Leu Lys Glu Val Asp Phe Gln Thr Leu Thr
275 280 285

Thr Pro Ala Pro Ala Pro Glu Glu Glu Asp Asp Asp Pro Asp Ala Pro
290 295 300

Asp Arg Pro Lys Lys Lys Lys Pro Lys Lys Asp Pro Leu Leu Ser Lys
305 310 315 320

Lys Ser Lys Ser Asp Pro Asn Ala Pro Ser Ile Asp Arg Ile Pro Leu
325 330 335

Pro Glu Leu Lys Asp Ser Asp Lys Leu Leu Lys Leu Lys Ala Leu Arg
340 345 350

Glu Ala Ser Lys Arg Leu Ala Leu Ser Lys Asp Gln Leu Pro Ser Ala
355 360 365

Val Phe Tyr Thr Val Leu Asn Ser His Gln Gly Val Thr Cys Ala Glu
370 375 380

Ile Ser Asp Asp Ser Thr Met Leu Ala Cys Gly Phe Gly Asp Ser Ser
385 390 395 400

Val Arg Ile Trp Ser Leu Thr Pro Ala Lys Leu Arg Thr Leu Lys Asp
405 410 415

Ala Asp Ser Leu Arg Glu Leu Asp Lys Glu Ser Ala Asp Ile Asn Val
420 425 430

Arg Met Leu Asp Asp Arg Ser Gly Glu Val Thr Arg Ser Leu Met Gly
435 440 445

His Thr Gly Pro Val Tyr Arg Cys Ala Phe Ala Pro Glu Met Asn Leu
450 455 460

Leu Leu Ser Cys Ser Glu Asp Ser Thr Ile Arg Leu Trp Ser Leu Leu
465 470 475 480

Thr Trp Ser Cys Val Val Thr Tyr Arg Gly His Val Tyr Pro Val Trp
485 490 495

Asp Val Arg Phe Ala Pro His Gly Tyr Tyr Phe Val Ser Cys Ser Tyr
500 505 510

Asp Lys Thr Ala Arg Leu Trp Ala Thr Asp Ser Asn Gln Ala Leu Arg
515 520 525

Val Phe Val Gly His Leu Ser Asp Val Asp Cys Val Gln Phe His Pro
530 535 540

Asn Ser Asn Tyr Val Ala Thr Gly Ser Ser Asp Arg Thr Val Arg Leu
545 550 555 560

Trp Asp Asn Met Thr Gly Gln Ser Val Arg Leu Met Thr Gly His Lys
565 570 575

Gly Ser Val Ser Ser Leu Ala Phe Ser Ala Cys Gly Arg Tyr Leu Ala
580 585 590

Ser Gly Ser Val Asp His Asn Ile Ile Ile Trp Asp Leu Ser Asn Gly
595 600 605

Ser Leu Val Thr Thr Leu Leu Arg His Thr Ser Thr Val Thr Thr Ile
610 615 620

Thr Phe Ser Arg Asp Gly Thr Val Leu Ala Ala Ala Gly Leu Asp Asn
625 630 635 640

Asn Leu Thr Leu Trp Asp Phe His Lys Val Thr Glu Asp Tyr Ile Ser
645 650 655

Asn His Ile Thr Val Ser His His Gln Asp Glu Asn Asp Glu Asp Val
660 665 670

Tyr Leu Met Arg Thr Phe Pro Ser Lys Asn Ser Pro Phe Val Ser Leu
675 680 685

His Phe Thr Arg Arg Asn Leu Leu Met Cys Val Gly Leu Phe Lys Ser
690 695 700

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2018 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 70..1842

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
GGAATTCGAG TTGGCCAAAG TGGCGCAATC CGGTATCAAT TGTTCAAACC
GAGCAGCCCC
    60

TCCAGCAGC ATG CTG TAC GGC TCC AGC ATC TCG GCG GAG TCC ATG
AAG
   108
          Met Leu Tyr Gly Ser Ser Ile Ser Ala Glu Ser Met
Lys
            1               5                  10


GTG ATC GCG GAG AGC ATC GGA GTG GGC TCC CTG TCG GAT GAC GCC
GCC
   156
Val Ile Ala Glu Ser Ile Gly Val Gly Ser Leu Ser Asp Asp Ala
Ala
     15                  20                  25

AAG GAA CTA GCG GAG GAT GTG TCC ATC AAG CTG AAG AGG ATT GTA
CAG
   204
Lys Glu Leu Ala Glu Asp Val Ser Ile Lys Leu Lys Arg Ile Val
Gln
 30                  35                  40
   45

GAT GCG GCC AAG TTC ATG AAC CAC GCC AAG CGG CAG AAG CTC TCA
GTG
   252
Asp Ala Ala Lys Phe Met Asn His Ala Lys Arg Gln Lys Leu Ser
Val
                50                  55
60

CGG GAC ATC GAC ATG TCC CTT AAG GTG CGA AAT GTG GAG CCG CAG
TAC
   300
Arg Asp Ile Asp Met Ser Leu Lys Val Arg Asn Val Glu Pro Gln
Tyr
            65                  70                  75
```

```
GGT TTC GTA GCC AAG GAC TTC ATT CCA CTC CGC TTC GCA TCT GGC
GGA
   348
Gly Phe Val Ala Lys Asp Phe Ile Pro Leu Arg Phe Ala Ser Gly
Gly
         80                  85                  90


GGA CGG GAG CTG CAC TTC ACC GAG GAC AAG GAA ATC GAC CTA GGA
GAA
   396
Gly Arg Glu Leu His Phe Thr Glu Asp Lys Glu Ile Asp Leu Gly
Glu
     95                  100                 105


ATC ACA TCC ACC AAC TCT GTA AAA ATT CCC CTG GAT CTC ACC CTG
CGC
   444
Ile Thr Ser Thr Asn Ser Val Lys Ile Pro Leu Asp Leu Thr Leu
Arg
110                 115                 120
  125

TCC CAT TGG TTT GTT GTG GAG GGA GTG CAA CCC ACT GTG CCC GAA
AAC
   492
Ser His Trp Phe Val Val Glu Gly Val Gln Pro Thr Val Pro Glu
Asn
                130                 135                 140

CCC CCT CCG CTC TCG AAG GAT TCC CAG TTA CTG GAC TCG GTC AAT
CCA
   540
Pro Pro Pro Leu Ser Lys Asp Ser Gln Leu Leu Asp Ser Val Asn
Pro
            145                 150                 155

GTT ATT AAG ATG GAT CAA GGC CTA AAC AAA GAT GCG GCA GGC AAA
CCC
   588
Val Ile Lys Met Asp Gln Gly Leu Asn Lys Asp Ala Ala Gly Lys
Pro
        160                 165                 170

ACC ACC GGC AAG ATA CAC AAG CTG AAA AAC GTG GAG ACC ATT CAT
GTC
   636
Thr Thr Gly Lys Ile His Lys Leu Lys Asn Val Glu Thr Ile His
Val
```

175 180 185

AAG CAA CTG GCC ACG CAC GAG TTG TCC GTG GAG CAG CAG TTG TAC TAC
684
Lys Gln Leu Ala Thr His Glu Leu Ser Val Glu Gln Gln Leu Tyr Tyr
190 195 200 205

AAG GAG ATC ACC GAG GCG TGC GTG GGA TCT GAT GAG CCG CGG CGC GGG
732
Lys Glu Ile Thr Glu Ala Cys Val Gly Ser Asp Glu Pro Arg Arg Gly
210 215 220

GAA GCG CTG CAG TCG CTG GGA TCC GAT CCT GGC CTG CAC GAA ATG CTT
780
Glu Ala Leu Gln Ser Leu Gly Ser Asp Pro Gly Leu His Glu Met Leu
225 230 235

CCC CGC ATG TGC ACC TTC ATT GCC GAG GGA GTT AAG GTC AAT GTG GTT
828
Pro Arg Met Cys Thr Phe Ile Ala Glu Gly Val Lys Val Asn Val Val
240 245 250

CAG AAC AAC TTG GCG TTG CTT ATT TAC CTC ATG CGC ATG GTT CGT GCG
876
Gln Asn Asn Leu Ala Leu Leu Ile Tyr Leu Met Arg Met Val Arg Ala
255 260 265

CTT CTG GAT AAT CCT TCG CTG TTT CTG GAG AAA TAC CTC CAC GAA CTG
924
Leu Leu Asp Asn Pro Ser Leu Phe Leu Glu Lys Tyr Leu His Glu Leu
270 275 280 285

ATA CCC TCG GTG ATG ACG TGC ATT GTG TCC AAA CAG CTG TGT ATG CGC
972

```
Ile Pro Ser Val Met Thr Cys Ile Val Ser Lys Gln Leu Cys Met
Arg
                290                 295                 300

CCC GAG CTG GAC AAT CAC TGG GCC CTG CGA GAC TTT GCC TCC CGA
CTG
  1020
Pro Glu Leu Asp Asn His Trp Ala Leu Arg Asp Phe Ala Ser Arg
Leu
            305                 310                 315

ATG GCT CAA ATC TGC AAG AAC TTC AAT ACC CTA ACC AAC AAT CTG
CAA
  1068
Met Ala Gln Ile Cys Lys Asn Phe Asn Thr Leu Thr Asn Asn Leu
Gln
        320                 325                 330

ACC CGT GTC ACC CGC ATC TTC AGC AAG GCC CTG CAG AAC GAC AAG
ACC
  1116
Thr Arg Val Thr Arg Ile Phe Ser Lys Ala Leu Gln Asn Asp Lys
Thr
    335                 340                 345

CAC CTG TCC TCG CTT TAC GGC TCT ATT GCG GGT CTC TCG GAG CTG
GGG
  1164
His Leu Ser Ser Leu Tyr Gly Ser Ile Ala Gly Leu Ser Glu Leu
Gly
350                 355                 360
  365

GGC GAA GTC ATA AAG GTT TTC ATC ATA CCC CGC CTT AAG TTC ATA
TCG
  1212
Gly Glu Val Ile Lys Val Phe Ile Ile Pro Arg Leu Lys Phe Ile
Ser
                370                 375                 380

GAG CGC ATT GAA CCT CAC CTG CTC GGC ACC TCC ATC AGC AAC ACT
GAC
  1260
Glu Arg Ile Glu Pro His Leu Leu Gly Thr Ser Ile Ser Asn Thr
Asp
            385                 390                 395
```

AAG ACA GCA GCA GGT CAC ATC CGC GCC ATG CTT CAG AAG TGC TGT CCC
1308
Lys Thr Ala Ala Gly His Ile Arg Ala Met Leu Gln Lys Cys Cys Pro
400 405 410

CCG ATT CTC AGG CAA ATG CTC AGC GCC AGA TAC AGC GGA GGA CTA CAA
1356
Pro Ile Leu Arg Gln Met Leu Ser Ala Arg Tyr Ser Gly Gly Leu Gln
415 420 425

GAA CGA CTT TGG CTT CCT GGG GCC GTC GCT GTG CCA GGC GTA GTC AAA
1404
Glu Arg Leu Trp Leu Pro Gly Ala Val Ala Val Pro Gly Val Val Lys
430 435 440
445

GTT CGA AAT GCG CCC GCC TCA AGC ATT GTA ACC CTG TCA TCC AAC ACT
1452
Val Arg Asn Ala Pro Ala Ser Ser Ile Val Thr Leu Ser Ser Asn Thr
450 455 460

ATC AAC ACG GCA CCC ATC ACG AGT GCA GCA CAA ACA GCA ACA ACC ATC
1500
Ile Asn Thr Ala Pro Ile Thr Ser Ala Ala Gln Thr Ala Thr Thr Ile
465 470 475

GGA CGA GTG TCC ATG CCC ACC ACA CAG AGA CAG GGA AGT CCC GGA GTC
1548
Gly Arg Val Ser Met Pro Thr Thr Gln Arg Gln Gly Ser Pro Gly Val
480 485 490

TCG TCC CTG CCG CAA ATA AGA GCC ATT CAG GCC AAC CAG CCG GCG CAA
1596
Ser Ser Leu Pro Gln Ile Arg Ala Ile Gln Ala Asn Gln Pro Ala Gln

```
     495                 500                 505

AAG TTT GTG ATA GTC ACC CAG AAC TCG CCG CAG CAG GGC CAG GCG
AAG
  1644
Lys Phe Val Ile Val Thr Gln Asn Ser Pro Gln Gln Gly Gln Ala
Lys
510                 515                 520
  525

GTG GTG CGG CGT GGC AGC TCT CCG CAC AGC GTG GTC CTC TCC GCG
GCC
  1692
Val Val Arg Arg Gly Ser Ser Pro His Ser Val Val Leu Ser Ala
Ala
                530                 535                 540

TCC AAC GCT GCC AGT GCC TCC AAT TCG AAC TCA AGC TCG AGC GGC
AGT
  1740
Ser Asn Ala Ala Ser Ala Ser Asn Ser Asn Ser Ser Ser Ser Gly
Ser
            545                 550                 555

CTA CTA GCG GCT GCA CAG CGG AGC AGC GAG AAT GTG TGT GTT ATT
GCC
  1788
Leu Leu Ala Ala Ala Gln Arg Ser Ser Glu Asn Val Cys Val Ile
Ala
        560                 565                 570

GGT AGC GAA GCG CCA GCA GTT GAT GGT ATA ACA GTT CAA TCT TTC
AGA
  1836
Gly Ser Glu Ala Pro Ala Val Asp Gly Ile Thr Val Gln Ser Phe
Arg
    575                 580                 585

GCA TCC TAGACGCCAA CTCGCTGATC ATTGAGACGG AGATTGTGCG
CGCACCGGCC
  1892
Ala Ser

590

CGAGCTGGCG GATCTCTCGC ACCTGGAGTA GCCAGCTTAG TTCGTAGTCC
ACATTTTGTC
  1952
```

ATATTGTATG CAATAAAATA AAAAATGCGG GTTCCTACCC CAAAAAAATG TAAAAAAAAA
2012

AAAAAA

2018

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 591 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Leu Tyr Gly Ser Ser Ile Ser Ala Glu Ser Met Lys Val Ile Ala
1 5 10 15

Glu Ser Ile Gly Val Gly Ser Leu Ser Asp Asp Ala Ala Lys Glu Leu
20 25 30

Ala Glu Asp Val Ser Ile Lys Leu Lys Arg Ile Val Gln Asp Ala Ala
35 40 45

Lys Phe Met Asn His Ala Lys Arg Gln Lys Leu Ser Val Arg Asp Ile
50 55 60

Asp Met Ser Leu Lys Val Arg Asn Val Glu Pro Gln Tyr Gly Phe Val
65 70 75 80

Ala Lys Asp Phe Ile Pro Leu Arg Phe Ala Ser Gly Gly Gly Arg Glu
85 90 95

Leu His Phe Thr Glu Asp Lys Glu Ile Asp Leu Gly Glu Ile Thr Ser
100 105 110

Thr Asn Ser Val Lys Ile Pro Leu Asp Leu Thr Leu Arg Ser His Trp
115 120 125

Phe Val Val Glu Gly Val Gln Pro Thr Val Pro Glu Asn Pro Pro Pro
130 135 140

Leu Ser Lys Asp Ser Gln Leu Leu Asp Ser Val Asn Pro Val Ile Lys
145 150 155 160

Met Asp Gln Gly Leu Asn Lys Asp Ala Ala Gly Lys Pro Thr Thr Gly
165 170 175

Lys Ile His Lys Leu Lys Asn Val Glu Thr Ile His Val Lys Gln Leu
180 185 190

Ala Thr His Glu Leu Ser Val Glu Gln Gln Leu Tyr Tyr Lys Glu Ile
195 200 205

Thr Glu Ala Cys Val Gly Ser Asp Glu Pro Arg Arg Gly Glu Ala Leu
210 215 220

Gln Ser Leu Gly Ser Asp Pro Gly Leu His Glu Met Leu Pro Arg Met
225 230 235 240

Cys Thr Phe Ile Ala Glu Gly Val Lys Val Asn Val Val Gln Asn Asn
245 250 255

Leu Ala Leu Leu Ile Tyr Leu Met Arg Met Val Arg Ala Leu Leu Asp
260 265 270

Asn Pro Ser Leu Phe Leu Glu Lys Tyr Leu His Glu Leu Ile Pro Ser
275 280 285

Val Met Thr Cys Ile Val Ser Lys Gln Leu Cys Met Arg Pro Glu Leu
290 295 300

Asp Asn His Trp Ala Leu Arg Asp Phe Ala Ser Arg Leu Met Ala Gln
305 310 315 320

Ile Cys Lys Asn Phe Asn Thr Leu Thr Asn Asn Leu Gln Thr Arg Val
325 330 335

Thr Arg Ile Phe Ser Lys Ala Leu Gln Asn Asp Lys Thr His Leu Ser
340 345 350

Ser Leu Tyr Gly Ser Ile Ala Gly Leu Ser Glu Leu Gly Gly Glu Val
355 360 365

Ile Lys Val Phe Ile Ile Pro Arg Leu Lys Phe Ile Ser Glu Arg Ile
370 375 380

Glu Pro His Leu Leu Gly Thr Ser Ile Ser Asn Thr Asp Lys Thr Ala
385 390 395 400

Ala Gly His Ile Arg Ala Met Leu Gln Lys Cys Cys Pro Pro Ile Leu
405 410 415

Arg Gln Met Leu Ser Ala Arg Tyr Ser Gly Gly Leu Gln Glu Arg Leu
420 425 430

Trp Leu Pro Gly Ala Val Ala Val Pro Gly Val Val Lys Val Arg Asn
435 440 445

Ala Pro Ala Ser Ser Ile Val Thr Leu Ser Ser Asn Thr Ile Asn Thr
450 455 460

Ala Pro Ile Thr Ser Ala Ala Gln Thr Ala Thr Thr Ile Gly Arg Val
465 470 475 480

Ser Met Pro Thr Thr Gln Arg Gln Gly Ser Pro Gly Val Ser Ser Leu
485 490 495

Pro Gln Ile Arg Ala Ile Gln Ala Asn Gln Pro Ala Gln Lys Phe Val
500 505 510

Ile Val Thr Gln Asn Ser Pro Gln Gln Gly Gln Ala Lys Val Val Arg
515 520 525

Arg Gly Ser Ser Pro His Ser Val Val Leu Ser Ala Ala Ser Asn Ala
530 535 540

Ala Ser Ala Ser Asn Ser Asn Ser Ser Ser Ser Gly Ser Leu Leu Ala
545 550 555 560

Ala Ala Gln Arg Ser Ser Glu Asn Val Cys Val Ile Ala Gly Ser Glu
565 570 575

Ala Pro Ala Val Asp Gly Ile Thr Val Gln Ser Phe Arg Ala Ser
580 585 590

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1120 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 80..913

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATATGTACG TGCACAATTT CAATGGAATA AACAATCTTC TTGCAGCAAA GCCGACGTAA
60

ACATAATAAC TATAGAAGT ATG AGC GCA GAG AAG TCC GAT AAG GCC AAG ATC
112
Met Ser Ala Glu Lys Ser Asp Lys Ala Lys Ile
1 5 10

AGT GCC CAA ATC AAG CAC GTG CCG AAG GAC GCG CAG GTG ATC ATG TCC
160
Ser Ala Gln Ile Lys His Val Pro Lys Asp Ala Gln Val Ile Met Ser
15 20 25

ATC CTG AAG GAG CTG AAT GTC CAG GAG TAC GAG CCG CGC GTG GTC AAC
208
Ile Leu Lys Glu Leu Asn Val Gln Glu Tyr Glu Pro Arg Val Val Asn
30 35 40

CAA CTG CTG GAG TTC ACC TTC CGC TAT GTC ACC TGC ATT CTG GAC GAC
256
Gln Leu Leu Glu Phe Thr Phe Arg Tyr Val Thr Cys Ile Leu Asp Asp
45 50 55

GCC AAG GTA TAC GCC AAC CAT GCG CGC AAG AAG ACC ATC GAC TTG GAC
304
Ala Lys Val Tyr Ala Asn His Ala Arg Lys Lys Thr Ile Asp Leu Asp
60 65 70
75

GAC GTG CGT CTG GCC ACC GAG GTT ACG CTG GAC AAG AGC TTC ACC GGG
352
Asp Val Arg Leu Ala Thr Glu Val Thr Leu Asp Lys Ser Phe Thr Gly

```
                    80                  85
90

CCG TTG GAG CGC CAC GTT CTA GCC AAG GTG GCC GAC GTG CGC AAC
AGC
   400
Pro Leu Glu Arg His Val Leu Ala Lys Val Ala Asp Val Arg Asn
Ser
            95                  100                 105

ATG CCC CTG CCA CCC ATT AAG CCG CAC TGC GGT CTC CGA CTG CCG
CCC
   448
Met Pro Leu Pro Pro Ile Lys Pro His Cys Gly Leu Arg Leu Pro
Pro
        110                 115                 120

GAC CGC TAC TGT CTC ACC GGC GTC AAC TAC AAA CTG CGG GCC ACT
AAT
   496
Asp Arg Tyr Cys Leu Thr Gly Val Asn Tyr Lys Leu Arg Ala Thr
Asn
    125                 130                 135

CAG CCC AAG AAA ATG ACC AAG TCG GCG GTG GAG GGC CGT CCA CTG
AAG
   544
Gln Pro Lys Lys Met Thr Lys Ser Ala Val Glu Gly Arg Pro Leu
Lys
140                 145                 150
  155

ACC GTC GTT AAG CCC GTC TCC AGC GCC AAT GGT CCG AAG AGG CCA
CAC
   592
Thr Val Val Lys Pro Val Ser Ser Ala Asn Gly Pro Lys Arg Pro
His
                160                 165                 170

TCC GTG GTG GCC AAG CAG CAG GTG GTG ACC ATT CCC AAG CCC GTC
ATC
   640
Ser Val Val Ala Lys Gln Gln Val Val Thr Ile Pro Lys Pro Val
Ile
            175                 180                 185

AAG TTT ACC ACC ACT ACG ACA ACG AAA ACG GTG GGC AGC TCC GGC
GGA
   688
```

```
Lys Phe Thr Thr Thr Thr Thr Lys Thr Val Gly Ser Ser Gly
Gly
        190                 195                 200


TCT GGG GGC GGC GGT GGT CAG GAG GTT AAG AGC GAG AGC ACC GGC
GCC
   736
Ser Gly Gly Gly Gly Gly Gln Glu Val Lys Ser Glu Ser Thr Gly
Ala
    205                 210                 215


GGC GGA GAT CTC AAG ATG GAG GTG GAC AGC GAT GCG GCG GCC GTG
GGC
   784
Gly Gly Asp Leu Lys Met Glu Val Asp Ser Asp Ala Ala Ala Val
Gly
220                 225                 230
  235

AGC ATC GCT GGC GCA TCC GGT TCG GGA GCA GGA AGT GCC AGC GGA
GGA
   832
Ser Ile Ala Gly Ala Ser Gly Ser Gly Ala Gly Ser Ala Ser Gly
Gly
                240                 245                 250

GGA GGA GGA GGA GGA TCA TCT GGC GTT GGA GTG GCC GTC AAG CGG
GAA
   880
Gly Gly Gly Gly Gly Ser Ser Gly Val Gly Val Ala Val Lys Arg
Glu
            255                 260                 265

CGT GAG GAG GAG GAG TTT GAG TTT GTG ACC AAC TAGCGAAACG
ACATCATTTA
   933
Arg Glu Glu Glu Glu Phe Glu Phe Val Thr Asn

        270                 275


CCTTAAATTA ATATTCTTAA ATCAGACCAA AGCACTTGCA TTTGGTTGAG
CGAACTGGGG
   993

GTCTAAATTT CAACTCGAAT GTGAAGTCCC AAAAACCTTA GTATAGATTC
GCCCGTTAAT
  1053
```

CATTATGAAA TCTACGTTTT ATACACAAAT ACAACTACCA GATTTTCATA TTAAAAAAAA
1113

AAAAAAA

1120

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ser Ala Glu Lys Ser Asp Lys Ala Lys Ile Ser Ala Gln Ile Lys
1 5 10 15

His Val Pro Lys Asp Ala Gln Val Ile Met Ser Ile Leu Lys Glu Leu
20 25 30

Asn Val Gln Glu Tyr Glu Pro Arg Val Val Asn Gln Leu Leu Glu Phe
35 40 45

Thr Phe Arg Tyr Val Thr Cys Ile Leu Asp Asp Ala Lys Val Tyr Ala
50 55 60

Asn His Ala Arg Lys Lys Thr Ile Asp Leu Asp Asp Val Arg Leu Ala
65 70 75 80

Thr Glu Val Thr Leu Asp Lys Ser Phe Thr Gly Pro Leu Glu Arg His
85 90 95

Val Leu Ala Lys Val Ala Asp Val Arg Asn Ser Met Pro Leu Pro Pro
100 105 110

Ile Lys Pro His Cys Gly Leu Arg Leu Pro Pro Asp Arg Tyr Cys Leu
115 120 125

Thr Gly Val Asn Tyr Lys Leu Arg Ala Thr Asn Gln Pro Lys Lys Met
130 135 140

Thr Lys Ser Ala Val Glu Gly Arg Pro Leu Lys Thr Val Val Lys Pro
145 150 155 160

Val Ser Ser Ala Asn Gly Pro Lys Arg Pro His Ser Val Val Ala Lys
165 170 175

Gln Gln Val Val Thr Ile Pro Lys Pro Val Ile Lys Phe Thr Thr Thr
180 185 190

Thr Thr Thr Lys Thr Val Gly Ser Ser Gly Gly Ser Gly Gly Gly Gly
195 200 205

Gly Gln Glu Val Lys Ser Glu Ser Thr Gly Ala Gly Gly Asp Leu Lys
210 215 220

Met Glu Val Asp Ser Asp Ala Ala Ala Val Gly Ser Ile Ala Gly Ala
225 230 235 240

Ser Gly Ser Gly Ala Gly Ser Ala Ser Gly Gly Gly Gly Gly Gly Gly
245 250 255

Ser Ser Gly Val Gly Val Ala Val Lys Arg Glu Arg Glu Glu Glu Glu
260 265 270

Phe Glu Phe Val Thr Asn
275

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5962 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 14..5692

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
TTATTTCCGG CAT ATG GGA CCC GGC TGC GAT TTG CTG CTG CGG ACA
GCA
    49
                Met Gly Pro Gly Cys Asp Leu Leu Leu Arg Thr
Ala
                  1               5                  10


GCT ACC ATC ACT GCT GCC GCC ATC ATG TCA GAC ACG GAC AGC GAC
GAA
    97
Ala Thr Ile Thr Ala Ala Ala Ile Met Ser Asp Thr Asp Ser Asp
Glu
         15                  20                  25


GAT TCC GCT GGA GGC GGC CCA TTT TCT TTA GCG GGT TTC CTT TTC
GGC
   145
Asp Ser Ala Gly Gly Gly Pro Phe Ser Leu Ala Gly Phe Leu Phe
Gly
     30                  35                  40


AAC ATC AAT GGA GCC GGG CAG CTG GAG GGG GAA AGC GTC TTG GAT
GAT
   193
Asn Ile Asn Gly Ala Gly Gln Leu Glu Gly Glu Ser Val Leu Asp
Asp
 45                  50                  55
   60

GAA TGT AAG AAG CAC TTG GCA GGC TTG GGG GCT TTG GGG CTG GGC
AGC
   241
Glu Cys Lys Lys His Leu Ala Gly Leu Gly Ala Leu Gly Leu Gly
Ser
```

65 70
75

CTG ATC ACT GAA CTC ACG GCA AAT GAA GAA TTG ACC GGG ACT GAC
GGT
289
Leu Ile Thr Glu Leu Thr Ala Asn Glu Glu Leu Thr Gly Thr Asp
Gly
80 85 90

GCC TTG GTA AAT GAT GAA GGG TGG GTT AGG AGT ACA GAA GAT GCT
GTG
337
Ala Leu Val Asn Asp Glu Gly Trp Val Arg Ser Thr Glu Asp Ala
Val
95 100 105

GAC TAT TCA GAC ATC AAT GAG GTG GCA GAA GAT GAA AGC CGA AGA
TAC
385
Asp Tyr Ser Asp Ile Asn Glu Val Ala Glu Asp Glu Ser Arg Arg
Tyr
110 115 120

CAG CAG ACG ATG GGG AGC TTG CAG CCC CTT TGC CAC TCA GAT TAT
GAT
433
Gln Gln Thr Met Gly Ser Leu Gln Pro Leu Cys His Ser Asp Tyr
Asp
125 130 135
140

deletion in hTAF$_{II}$250Δ cloned by PCR

GAA GAT GAC TAT GAT [GCT GAT TGT GAA GAC ATT GAT TGC AAG TTG
ATG
481
Glu Asp Asp Tyr Asp Ala Asp Cys Glu Asp Ile Asp Cys Lys Leu
Met
145 150 155

CCT CCT CCA CCT CCA CCC CCG GGA CCA ATG AAG AAG GAT AAG GAC
CAG
529
Pro Pro Pro Pro Pro Pro Pro Gly Pro Met Lys Lys Asp Lys Asp
Gln
160 165 170

Insert in alternatively spliced hTAF$_{II}$250 cDNA clone

GAT TCT ATT ACT GGT {GTG TCT GAA AAT GGA GAA GGC ATC ATC TTG
CCC
577

```
Asp Ser Ile Thr Gly Val Ser Glu Asn Gly Glu Gly Ile Ile Leu
Pro
        175                 180                 185

TCC ATC ATT GCC CCT TCC TCT TTG GCC TCA GAG AAA GTG GAC TTC
AGT
   625
Ser Ile Ile Ala Pro Ser Ser Leu Ala Ser Glu Lys Val Asp Phe
Ser
    190                 195                 200

AGT TCC TCT GAC TCA GAA TCT GAG ATG GGA CCT CAG GAA GCA ACA
CAG
   673
Ser Ser Ser Asp Ser Glu Ser Glu Met Gly Pro Gln Glu Ala Thr
Gln
205                 210                 215
  220

GCA GAA TCT GAA GAT GGA AAG CTG ACC CTT CCA TTG GCT GGG ATT
ATG
   721
Ala Glu Ser Glu Asp Gly Lys Leu Thr Leu Pro Leu Ala Gly Ile
Met
               225                 230                 235

CAG CAT GAT GCC ACC AAG CTG TTG CCA AGT GTC ACA GAA CTT TTT
CCA
   769
Gln His Asp Ala Thr Lys Leu Leu Pro Ser Val Thr Glu Leu Phe
Pro
            240                 245                 250

GAA TTT CGA CCT GGA AAG GTG TTA CGT TTT CTA CGT CTT TTT GGA
CCA
   817
Glu Phe Arg Pro Gly Lys Val Leu Arg Phe Leu Arg Leu Phe Gly
Pro
        255                 260                 265

GGG AAG AAT GTC CCA TCT GTT TGG CGG AGT GCT CGG AGA AAG AGG
AAG
   865
Gly Lys Asn Val Pro Ser Val Trp Arg Ser Ala Arg Arg Lys Arg
Lys
    270                 275                 280
```

```
AAG AAG CAC CGT GAG CTG ATA CAG GAA GAG CAG ATC CAG GAG GTG
GAG
   913
Lys Lys His Arg Glu Leu Ile Gln Glu Glu Gln Ile Gln Glu Val
Glu
285                 290                 295
  300

TGC TCA GTA GAA TCA GAA GTC AGC CAG AAG TCT TTG TGG AAC TAC
GAC
   961
Cys Ser Val Glu Ser Glu Val Ser Gln Lys Ser Leu Trp Asn Tyr
Asp
               305                 310                 315

TAC GCT CCA CCA CCA CCT CCA GAG CAG TGT CTC TCT GAT GAT GAA
ATC
  1009
Tyr Ala Pro Pro Pro Pro Pro Glu Gln Cys Leu Ser Asp Asp Glu
Ile
           320                 325                 330

ACG ATG ATG GCT CCT GTG GAG TCC AAA TTT TCC CAA TCA ACT GGA
GAT
  1057
Thr Met Met Ala Pro Val Glu Ser Lys Phe Ser Gln Ser Thr Gly
Asp
        335                 340                 345

ATA GAT AAA GTG ACA GAT ACC AAA CCA AGA GTG GCT GAG TGG CGT
TAT
  1105
Ile Asp Lys Val Thr Asp Thr Lys Pro Arg Val Ala Glu Trp Arg
Tyr
    350                 355                 360

GGG CCT GCC CGA CTG TGG TAT GAT ATG CTG GGT GTC CCT GAA GAT
GGC
  1153
Gly Pro Ala Arg Leu Trp Tyr Asp Met Leu Gly Val Pro Glu Asp
Gly
365                 370                 375
  380

AGT GGG TTT GAC TAT GGC TTC AAA CTG AGA AAG ACA GAA CAT GAA
CCT
  1201
Ser Gly Phe Asp Tyr Gly Phe Lys Leu Arg Lys Thr Glu His Glu
Pro
```

385 390 395

GTG ATA AAA TCT AGA ATG ATA GAG GAA TTT AGG AAA CTT GAG GAA AAC
1249
Val Ile Lys Ser Arg Met Ile Glu Glu Phe Arg Lys Leu Glu Glu Asn
400 405 410

AAT GGC ACT GAT CTT CTG GCT GAT GAA AAC TTC CTG ATG GTG ACA CAG
1297
Asn Gly Thr Asp Leu Leu Ala Asp Glu Asn Phe Leu Met Val Thr Gln
415 420 425

CTG CAT TGG GAG GAT GAT ATC ATC TGG GAT GGG GAG GAT GTC AAA CAC
1345
Leu His Trp Glu Asp Asp Ile Ile Trp Asp Gly Glu Asp Val Lys His
430 435 440

AAA GGG ACA AAA CCT CAG CGT GCA AGC CTG GCA GGC TGG CTT CCT TCT
1393
Lys Gly Thr Lys Pro Gln Arg Ala Ser Leu Ala Gly Trp Leu Pro Ser
445 450 455
460

AGC ATG ACT AGG AAT GCG ATG GCT TAC AAT GTT CAG CAA GGT TTT GCA
1441
Ser Met Thr Arg Asn Ala Met Ala Tyr Asn Val Gln Gln Gly Phe Ala
465 470 475

GCC ACT CTT GAT GAT GAC AAA CCT TGG TAC TCC ATT TTT CCC ATT GAC
1489
Ala Thr Leu Asp Asp Asp Lys Pro Trp Tyr Ser Ile Phe Pro Ile Asp
480 485 490

AAT GAG GAT CTG GTA TAT GGA CGC TGG GAG GAC AAT ATC ATT TGG GAT
1537

```
Asn Glu Asp Leu Val Tyr Gly Arg Trp Glu Asp Asn Ile Ile Trp
Asp
        495                 500                 505


GCT CAG GCC ATG CCC CGG CTG TTG GAA CCT CCT GTT TTG ACA CTT
GAT
  1585
Ala Gln Ala Met Pro Arg Leu Leu Glu Pro Pro Val Leu Thr Leu
Asp
    510                 515                 520


CCC AAT GAT GAG AAC CTC ATT TTG GAA ATT CCT GAT GAG AAG GAA
GAG
  1633
Pro Asn Asp Glu Asn Leu Ile Leu Glu Ile Pro Asp Glu Lys Glu
Glu
525                 530                 535
  540

GCC ACC TCT AAC TCC CCC TCC AAG GAG AGT AAG AAG GAA TCA TCT
CTG
  1681
Ala Thr Ser Asn Ser Pro Ser Lys Glu Ser Lys Lys Glu Ser Ser
Leu
                545                 550                 555

AAG AAG AGT CGA ATT CTC TTA GGG AAA ACA GGA GTC ATC AAG GAG
GAA
  1729
Lys Lys Ser Arg Ile Leu Leu Gly Lys Thr Gly Val Ile Lys Glu
Glu
            560                 565                 570

CCA CAG CAG AAC ATG TCT CAG CCA GAA GTG AAA GAT CCA TGG AAT
CTC
  1777
Pro Gln Gln Asn Met Ser Gln Pro Glu Val Lys Asp Pro Trp Asn
Leu
        575                 580                 585

TCC AAT GAT GAG TAT TAT TAT CCC AAG CAA CAG GGT CTT CGA GGC
ACC
  1825
Ser Asn Asp Glu Tyr Tyr Tyr Pro Lys Gln Gln Gly Leu Arg Gly
Thr
    590                 595                 600
```

```
TTT GGA GGG AAT ATT ATC CAG CAT TCA ATT CCT GCT GTG GAA TTA
CGG
  1873
Phe Gly Gly Asn Ile Ile Gln His Ser Ile Pro Ala Val Glu Leu
Arg
605                 610                 615
  620

CAG CCC TTC TTT CCC ACC CAC ATG GGG CCC ATC AAA CTC CGG CAG
TTC
  1921
Gln Pro Phe Phe Pro Thr His Met Gly Pro Ile Lys Leu Arg Gln
Phe
                625                 630                 635

CAT CGC CCA CCT CTG AAA AAG TAC TCA TTT GGT GCA CTT TCT CAG
CCA
  1969
His Arg Pro Pro Leu Lys Lys Tyr Ser Phe Gly Ala Leu Ser Gln
Pro
            640                 645                 650

GGT CCC CAC TCA GTC CAA CCT TTG CTA AAG CAC ATC AAA AAA AAG
GCC
  2017
Gly Pro His Ser Val Gln Pro Leu Leu Lys His Ile Lys Lys Lys
Ala
        655                 660                 665

AAG ATG AGA GAA CAA GAG AGG CAA GCT TCA GGT GGT GGA GAG ATG
TTT
  2065
Lys Met Arg Glu Gln Glu Arg Gln Ala Ser Gly Gly Gly Glu Met
Phe
    670                 675                 680

TTT ATG CGC ACA CCT CAG GAC CTC ACA GGC AAA GAT GGT GAT CTT
ATT
  2113
Phe Met Arg Thr Pro Gln Asp Leu Thr Gly Lys Asp Gly Asp Leu
Ile
685                 690                 695
  700

CTT GCA GAA TAT AGT GAG GAA AAT GGA CCC TTA ATG ATG CAG GTT
GGC
  2161
Leu Ala Glu Tyr Ser Glu Glu Asn Gly Pro Leu Met Met Gln Val
Gly
```

705 710 715

ATG GCA ACC AAG ATA AAG AAC TAT TAT AAA CGG AAA CCT GGA AAA GAT
2209
Met Ala Thr Lys Ile Lys Asn Tyr Tyr Lys Arg Lys Pro Gly Lys Asp
720 725 730

CCT GGA GCA CCA GAT TGT AAA TAT GGG GAA ACT GTT TAC TGC CAT ACA
2257
Pro Gly Ala Pro Asp Cys Lys Tyr Gly Glu Thr Val Tyr Cys His Thr
735 740 745

TCT CCT TTC CTG GGT TCT CTC CAT CCT GGC CAA TTG CTG CAA GCA TTT
2305
Ser Pro Phe Leu Gly Ser Leu His Pro Gly Gln Leu Leu Gln Ala Phe
750 755 760

GAG AAC AAC CTT TTT CGT GCT CCA ATT TAT CTT CAT AAG ATG CCA GAA
2353
Glu Asn Asn Leu Phe Arg Ala Pro Ile Tyr Leu His Lys Met Pro Glu
765 770 775
780

ACT GAT TTC TTG ATC ATT CGG ACA AGA CAG GGT TAC TAT ATT CGG GAA
2401
Thr Asp Phe Leu Ile Ile Arg Thr Arg Gln Gly Tyr Tyr Ile Arg Glu
785 790 795

TTA GTG GAT ATT TTT GTG GTT GGC CAG CAG TGT CCC TTG TTT GAA GTT
2449
Leu Val Asp Ile Phe Val Val Gly Gln Gln Cys Pro Leu Phe Glu Val
800 805 810

CCT GGG CCT AAC TCC AAA AGG GCC AAT ACG CAT ATT CGA GAC TTT CTA
2497

Pro Gly Pro Asn Ser Lys Arg Ala Asn Thr His Ile Arg Asp Phe Leu
815 820 825

CAG GTT TTT ATT TAC CGC CTT TTC TGG AAA AGT AAA GAT CGG CCA CGG
2545
Gln Val Phe Ile Tyr Arg Leu Phe Trp Lys Ser Lys Asp Arg Pro Arg
830 835 840

AGG ATA CGA ATG GAA GAT ATA AAA AAA GCC TTT CCT TCC CAT TCA GAA
2593
Arg Ile Arg Met Glu Asp Ile Lys Lys Ala Phe Pro Ser His Ser Glu
845 850 855 860

AGC AGC ATC CGG AAG AGG CTA AAG CTC TGC GCT GAC TTC AAA CGC ACA
2641
Ser Ser Ile Arg Lys Arg Leu Lys Leu Cys Ala Asp Phe Lys Arg Thr
865 870 875

GGG ATG GAC TCA AAC TGG TGG GTG CTT AAG TCT GAT TTT CGT TTA CCA
2689
Gly Met Asp Ser Asn Trp Trp Val Leu Lys Ser Asp Phe Arg Leu Pro
880 885 890

ACG GAA GAA GAG ATC AGA GCT ATG GTG TCA CCA GAG CAG TGC TGT GCT
2737
Thr Glu Glu Glu Ile Arg Ala Met Val Ser Pro Glu Gln Cys Cys Ala
895 900 905

TAT TAT AGC ATG ATA GCT GCA GAG CAA CGA CTG AAG GAT GCT GGC TAT
2785
Tyr Tyr Ser Met Ile Ala Ala Glu Gln Arg Leu Lys Asp Ala Gly Tyr
910 915 920

```
GGT GAG AAA TCC TTT TTT GCT CCA GAA GAA GAA AAT GAG GAA GAT
TTC
  2833
Gly Glu Lys Ser Phe Phe Ala Pro Glu Glu Glu Asn Glu Glu Asp
Phe
925                 930                 935
  940

CAG ATG AAG ATT GAT GAT GAA GTT CGC ACT GCC CCT TGG AAC ACC
ACA
  2881
Gln Met Lys Ile Asp Asp Glu Val Arg Thr Ala Pro Trp Asn Thr
Thr
                945                 950                 955

AGG GCC TTC ATT GCT GCC ATG AAG GGC AAG TGT CTG CTA GAG GTG
ACT
  2929
Arg Ala Phe Ile Ala Ala Met Lys Gly Lys Cys Leu Leu Glu Val
Thr
            960                 965                 970

GGG GTG GCA GAT CCC ACG GGG TGT GGT GAA GGA TTC TCC TAT GTG
AAG
  2977
Gly Val Ala Asp Pro Thr Gly Cys Gly Glu Gly Phe Ser Tyr Val
Lys
        975                 980                 985

ATT CCA AAC AAA CCA ACA CAG CAG AAG GAT GAT AAA GAA CCG CAG
CCA
  3025
Ile Pro Asn Lys Pro Thr Gln Gln Lys Asp Asp Lys Glu Pro Gln
Pro
    990                 995                 1000

GTG AAG AAG ACA GTG ACA GGA ACA GAT GCA GAC CTT CGT CGC CTT
TCC
  3073
Val Lys Lys Thr Val Thr Gly Thr Asp Ala Asp Leu Arg Arg Leu
Ser
1005                1010                1015
  1020

CTG AAA AAT GCC AAG CAA CTT CTA CGT AAA TTT GGT GTG CCT GAG
GAA
  3121
Leu Lys Asn Ala Lys Gln Leu Leu Arg Lys Phe Gly Val Pro Glu
Glu
```

```
                1025                1030                1035

GAG ATT AAA AAG TTG TCC CGC TGG GAA GTG ATT GAT GTG GTG CGC
ACA
  3169
Glu Ile Lys Lys Leu Ser Arg Trp Glu Val Ile Asp Val Val Arg
Thr
            1040                1045                1050

ATG TCA ACA GAA CAG GCT CGT TCT GGA GAG GGG CCC ATG AGT AAA
TTT
  3217
Met Ser Thr Glu Gln Ala Arg Ser Gly Glu Gly Pro Met Ser Lys
Phe
        1055                1060                1065

GCC CGT GGA TCA AGG TTT TCT GTG GCT GAG CAT CAA GAG CGT TAC
AAA
  3265
Ala Arg Gly Ser Arg Phe Ser Val Ala Glu His Gln Glu Arg Tyr
Lys
    1070                1075                1080

GAG GAA TGT CAG CGC ATC TTT GAC CTA CAG AAC AAG GTT CTG TCA
TCA
  3313
Glu Glu Cys Gln Arg Ile Phe Asp Leu Gln Asn Lys Val Leu Ser
Ser
1085                1090                1095
  1100

ACT GAA GTC TTA TCA ACT GAC ACA GAC AGC AGC TCA GCT GAA GAT
AGT
  3361
Thr Glu Val Leu Ser Thr Asp Thr Asp Ser Ser Ser Ala Glu Asp
Ser
            1105                1110                1115

GAC TTT GAA GAA ATG GGA AAG AAC ATT GAG AAC ATG TTG CAG AAC
AAG
  3409
Asp Phe Glu Glu Met Gly Lys Asn Ile Glu Asn Met Leu Gln Asn
Lys
            1120                1125                1130

AAA ACC AGC TCT CAG CTT TCA CGT GAA CGG GAG GAA CAG GAG CGG
AAG
  3457
```

```
Lys Thr Ser Ser Gln Leu Ser Arg Glu Arg Glu Glu Gln Glu Arg
Lys
        1135                1140                1145


GAA CTA CAG CGA ATG CTA CTG GCA GCA GGC TCA GCA GCA TCC GGA
AAC
  3505
Glu Leu Gln Arg Met Leu Leu Ala Ala Gly Ser Ala Ala Ser Gly
Asn
    1150                1155                1160


AAT CAC AGA GAT GAT GAC ACA GCT TCC GTG ACT AGC CTT AAC TCT
TCT
  3553
Asn His Arg Asp Asp Asp Thr Ala Ser Val Thr Ser Leu Asn Ser
Ser
1165                1170                1175
  1180

GCC ACT GGA CGC TGT CTC AAG ATT TAT CGC ACG TTT CGA GAT GAA
GAG
  3601
Ala Thr Gly Arg Cys Leu Lys Ile Tyr Arg Thr Phe Arg Asp Glu
Glu
                1185                1190                1195

GGG AAA GAG TAT GTT CGC TGT GAG ACA GTC CGA AAA CCA GCT GTC
ATT
  3649
Gly Lys Glu Tyr Val Arg Cys Glu Thr Val Arg Lys Pro Ala Val
Ile
            1200                1205                1210

GAT GCC TAT GTG CGC ATA CGG ACT ACA AAA GAT GAG GAA TTC ATT
CGA
  3697
Asp Ala Tyr Val Arg Ile Arg Thr Thr Lys Asp Glu Glu Phe Ile
Arg
        1215                1220                1225

AAA TTT GCC CTT TTT GAT GAA CAA CAT CGG GAA GAG ATG CGA AAA
GAA
  3745
Lys Phe Ala Leu Phe Asp Glu Gln His Arg Glu Glu Met Arg Lys
Glu
    1230                1235                1240
```

```
CGG CGG AGG ATT CAA GAG CAA CTG AGG CGG CTT AAG AGG AAC CAG
GAA
  3793
Arg Arg Arg Ile Gln Glu Gln Leu Arg Arg Leu Lys Arg Asn Gln
Glu
1245                1250                1255
  1260

AAG GAG AAG CTT AAG GGT CCT CCT GAG AAG AAG CCC AAG AAA ATG
AAG
  3841
Lys Glu Lys Leu Lys Gly Pro Pro Glu Lys Lys Pro Lys Lys Met
Lys
                1265                1270                1275

GAG CGT CCT GAC CTA AAA CTG AAA TGT GGG GCA TGT GGT GCC ATT
GGA
  3889
Glu Arg Pro Asp Leu Lys Leu Lys Cys Gly Ala Cys Gly Ala Ile
Gly
            1280                1285                1290

CAC ATG AGG ACT AAC AAA TTC TGC CCC CTC TAT TAT CAA ACA AAT
GCG
  3937
His Met Arg Thr Asn Lys Phe Cys Pro Leu Tyr Tyr Gln Thr Asn
Ala
        1295                1300                1305

CCA CCT TCC AAC CCT GTT GCC ATG ACA GAA GAA CAG GAG GAG GAG
TTG
  3985
Pro Pro Ser Asn Pro Val Ala Met Thr Glu Glu Gln Glu Glu Glu
Leu
    1310                1315                1320

GAA AAG ACA GTC ATT CAT AAT GAT AAT GAA GAA CTT ATC AAG GTT
GAA
  4033
Glu Lys Thr Val Ile His Asn Asp Asn Glu Glu Leu Ile Lys Val
Glu
1325                1330                1335
  1340

GGG ACC AAA ATT GTC TTG GGG AAA CAG CTA ATT GAG AGT GCG GAT
GAG
  4081
Gly Thr Lys Ile Val Leu Gly Lys Gln Leu Ile Glu Ser Ala Asp
Glu
```

1345 1350 1355

GTT CGC AGA AAA TCT CTG GTT CTC AAG TTT CCT AAA CAG CAG CTT CCT
4129
Val Arg Arg Lys Ser Leu Val Leu Lys Phe Pro Lys Gln Gln Leu Pro
1360 1365 1370

CCA AAG AAG AAA CGG CGA GTT GGA ACC ACT GTT CAC TGT GAC TAT TTG
4177
Pro Lys Lys Lys Arg Arg Val Gly Thr Thr Val His Cys Asp Tyr Leu
1375 1380 1385

AAT AGA CCT CAT AAG TCC ATC CAC CGG CGC CGC ACA GAC CCT ATG GTG
4225
Asn Arg Pro His Lys Ser Ile His Arg Arg Arg Thr Asp Pro Met Val
1390 1395 1400

ACG CTG TCG TCC ATC TTG GAG TCT ATC ATC AAT GAC ATG AGA GAT CTT
4273
Thr Leu Ser Ser Ile Leu Glu Ser Ile Ile Asn Asp Met Arg Asp Leu
1405 1410 1415
1420

CCA AAT ACA TAC CCT TTC CAC ACT CCA GTC AAT GCA AAG GTT GTA AAG
4321
Pro Asn Thr Tyr Pro Phe His Thr Pro Val Asn Ala Lys Val Val Lys
1425 1430 1435

GAC TAC TAC AAA ATC ATC ACT CGG CCA ATG GAC CTA CAA ACA CTC CGC
4369
Asp Tyr Tyr Lys Ile Ile Thr Arg Pro Met Asp Leu Gln Thr Leu Arg
1440 1445 1450

GAA AAC GTG CGT AAA CGC CTC TAC CCA TCT CGG GAA GAG TTC AGA GAG
4417

```
Glu Asn Val Arg Lys Arg Leu Tyr Pro Ser Arg Glu Glu Phe Arg
Glu
        1455                1460                1465

CAT CTG GAG CTA ATT GTG AAA AAT AGT GCA ACC TAC AAT GGG CCA
AAA
  4465
His Leu Glu Leu Ile Val Lys Asn Ser Ala Thr Tyr Asn Gly Pro
Lys
    1470                1475                1480

CAC TCA TTG ACT CAG ATC TCT CAA TCC ATG CTG GAT CTC TGT GAT
GAA
  4513
His Ser Leu Thr Gln Ile Ser Gln Ser Met Leu Asp Leu Cys Asp
Glu
1485                1490                1495
  1500

AAA CTC AAA GAG AAA GAA GAC AAA TTA GCT CGC TTA GAG AAA GCT
ATC
  4561
Lys Leu Lys Glu Lys Glu Asp Lys Leu Ala Arg Leu Glu Lys Ala
Ile
             1505                1510                1515

AAC CCC TTG CTG GAT GAT GAT GAC CAA GTG GCG TTT TCT TTC ATT
CTG
  4609
Asn Pro Leu Leu Asp Asp Asp Asp Gln Val Ala Phe Ser Phe Ile
Leu
         1520                1525                1530

GAC AAC ATT GTC ACC CAG AAA ATG ATG GCA GTT CCA GAT TCT TGG
CCA
  4657
Asp Asn Ile Val Thr Gln Lys Met Met Ala Val Pro Asp Ser Trp
Pro
        1535                1540                1545

TTT CAT CAC CCA GTT AAT AAG AAA TTT GTT CCA GAT TAT TAC AAA
GTG
  4705
Phe His His Pro Val Asn Lys Lys Phe Val Pro Asp Tyr Tyr Lys
Val
    1550                1555                1560
```

```
ATT GTC AAT CCA ATG GAT TTA GAG ACC ATA CGT AAG AAC ATC TCC
AAG
  4753
Ile Val Asn Pro Met Asp Leu Glu Thr Ile Arg Lys Asn Ile Ser
Lys
1565                1570                1575
  1580

CAC AAG TAT CAG AGT CGG GAG AGC TTT CTG GAT GAT GTA AAC CTT
ATT
  4801
His Lys Tyr Gln Ser Arg Glu Ser Phe Leu Asp Asp Val Asn Leu
Ile
               1585                1590                1595


CTG GCC AAC AGT GTT AAG TAT AAT GGA CCT GAG AGT CAG TAT ACT
AAG
  4849
Leu Ala Asn Ser Val Lys Tyr Asn Gly Pro Glu Ser Gln Tyr Thr
Lys
           1600                1605                1610


ACT GCC CAG GAG ATT GTG AAC GTC TGT TAC CAG ACA TTG ACT GAG
TAT
  4897
Thr Ala Gln Glu Ile Val Asn Val Cys Tyr Gln Thr Leu Thr Glu
Tyr
       1615                1620                1625


GAT GAA CAT TTG ACT CAA CTT GAG AAG GAT ATT TGT ACT GCT AAA
GAA
  4945
Asp Glu His Leu Thr Gln Leu Glu Lys Asp Ile Cys Thr Ala Lys
Glu
    1630                1635                1640


GCA GCT TTG GAG GAA GCA GAA TTA GAA AGC CTG GAC CCA ATG ACC
CCA
  4993
Ala Ala Leu Glu Glu Ala Glu Leu Glu Ser Leu Asp Pro Met Thr
Pro
1645                1650                1655
  1660

GGG CCC TAC ACG CCT CAG CCT CCT GAT TTG TAT GAT ACC AAC ACA
TCC
  5041
Gly Pro Tyr Thr Pro Gln Pro Pro Asp Leu Tyr Asp Thr Asn Thr
Ser
```

```
                    1665                1670                1675

CTC AGT ATG TCT CGA GAT GCC TCT GTA TTT CAA GAT GAG AGC AAT
ATG
  5089
Leu Ser Met Ser Arg Asp Ala Ser Val Phe Gln Asp Glu Ser Asn
Met
            1680                1685                1690

TCT GTC TTG GAT ATC CCC AGT GCC ACT CCA GAA AAG CAG GTA ACA
CAG
  5137
Ser Val Leu Asp Ile Pro Ser Ala Thr Pro Glu Lys Gln Val Thr
Gln
        1695                1700                1705

GAA GGT GAA GAT GGA GAT GGT GAT CTT GCA GAT GAA GAG GAA GGA
ACT
  5185
Glu Gly Glu Asp Gly Asp Gly Asp Leu Ala Asp Glu Glu Glu Gly
Thr
    1710                1715                1720

GTA CAA CAG CCT CAA GCC AGT GTC CTG TAT GAG GAT TTG CTT ATG
TCT
  5233
Val Gln Gln Pro Gln Ala Ser Val Leu Tyr Glu Asp Leu Leu Met
Ser
1725                1730                1735
  1740

GAA GGA GAA GAT GAT GAG GAA GAT GCT GGG AGT GAT GAA GAA GGA
GAC
  5281
Glu Gly Glu Asp Asp Glu Glu Asp Ala Gly Ser Asp Glu Glu Gly
Asp
                1745                1750                1755

AAT CCT TTC TCT GCT ATC CAG CTG AGT GAA AGT GGA AGT GAC TCT
GAT
  5329
Asn Pro Phe Ser Ala Ile Gln Leu Ser Glu Ser Gly Ser Asp Ser
Asp
            1760                1765                1770

GTG GGA TCT GGT GGA ATA AGA CCC AAA CAA CCC CGC ATG CTT CAG
GAG
  5377
```

```
Val Gly Ser Gly Gly Ile Arg Pro Lys Gln Pro Arg Met Leu Gln
Glu
        1775                1780                1785


AAC ACA AGG ATG GAC ATG GAA AAT GAA GAA AGC ATG ATG TCC TAT
GAG
  5425
Asn Thr Arg Met Asp Met Glu Asn Glu Glu Ser Met Met Ser Tyr
Glu
    1790                1795                1800


GGA GAC GGT GGG GAG GCT TCC CAT GGT TTG GAG GAT AGC AAC ATC
AGT
  5473
Gly Asp Gly Gly Glu Ala Ser His Gly Leu Glu Asp Ser Asn Ile
Ser
1805                1810                1815
  1820

TAT GGG AGC TAT GAG GAG CCT GAT CCC AAG TCG AAC ACC CAA GAC
ACA
  5521
Tyr Gly Ser Tyr Glu Glu Pro Asp Pro Lys Ser Asn Thr Gln Asp
Thr
                1825                1830                1835


AGC TTC AGC AGC ATC GGT GGG TAT GAG GTA TCA GAG GAG GAA GAA
GAT
  5569
Ser Phe Ser Ser Ile Gly Gly Tyr Glu Val Ser Glu Glu Glu Glu
Asp
            1840                1845                1850


GAG GAG GAG GAA GAG CAG CGC TCT GGG CCG AGC GTA CTA AGC CAG
GTC
  5617
Glu Glu Glu Glu Glu Gln Arg Ser Gly Pro Ser Val Leu Ser Gln
Val
        1855                1860                1865


CAC CTG TCA GAG GAC GAG GAG GAC AGT GAG GAT TTC CAC TCC ATT
GCT
  5665
His Leu Ser Glu Asp Glu Glu Asp Ser Glu Asp Phe His Ser Ile
Ala
    1870                1875                1880
```

GGG GAC AGT GAC TTG GAC TCT GAT GAA TGAGGCTTCC TTTGGGCCTC

5712
Gly Asp Ser Asp Leu Asp Ser Asp Glu

1885 1890

CTTGGTCAGC CTTCCCTGTT CTCCAGCCTA GGTGGTTCAC CTTTCCCCAA
TTTGTTCATA
5772

TTTGTACAGT ATCTGATCCT GAAATCATGA AATTAACTAA CACCTTAGCC
TTTTTAAAAG
5832

TAGTAAGTAA ATGATAATAA ATCACCTCTC CTAATCTTCC TGGGGCAATG
TCACCCTTTG
5892

ATTTAAAACA AAGCAACCCC CTTTCCCCTA CCACTACGGA AAAGAGCAAG
CTCATTTTTC
5952

CGTGTCCTCC

5962

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1893 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Gly Pro Gly Cys Asp Leu Leu Leu Arg Thr Ala Ala Thr Ile
Thr
1 5 10
15

Ala Ala Ala Ile Met Ser Asp Thr Asp Ser Asp Glu Asp Ser Ala
Gly
20 25 30

Gly Gly Pro Phe Ser Leu Ala Gly Phe Leu Phe Gly Asn Ile Asn
Gly
35 40 45

Ala Gly Gln Leu Glu Gly Glu Ser Val Leu Asp Asp Glu Cys Lys Lys
50 55 60

His Leu Ala Gly Leu Gly Ala Leu Gly Leu Gly Ser Leu Ile Thr Glu
65 70 75 80

Leu Thr Ala Asn Glu Glu Leu Thr Gly Thr Asp Gly Ala Leu Val Asn
85 90 95

Asp Glu Gly Trp Val Arg Ser Thr Glu Asp Ala Val Asp Tyr Ser Asp
100 105 110

Ile Asn Glu Val Ala Glu Asp Glu Ser Arg Arg Tyr Gln Gln Thr Met
115 120 125

Gly Ser Leu Gln Pro Leu Cys His Ser Asp Tyr Asp Glu Asp Asp Tyr
130 135 140

Asp Ala Asp Cys Glu Asp Ile Asp Cys Lys Leu Met Pro Pro Pro Pro
145 150 155 160

Pro Pro Pro Gly Pro Met Lys Lys Asp Lys Asp Gln Asp Ser Ile Thr
165 170 175

Gly Val Ser Glu Asn Gly Glu Gly Ile Ile Leu Pro Ser Ile Ile Ala
180 185 190

Pro Ser Ser Leu Ala Ser Glu Lys Val Asp Phe Ser Ser Ser Ser Asp
195 200 205

Ser Glu Ser Glu Met Gly Pro Gln Glu Ala Thr Gln Ala Glu Ser Glu
210 215 220

Asp Gly Lys Leu Thr Leu Pro Leu Ala Gly Ile Met Gln His Asp Ala
225 230 235 240

Thr Lys Leu Leu Pro Ser Val Thr Glu Leu Phe Pro Glu Phe Arg Pro
245 250 255

Gly Lys Val Leu Arg Phe Leu Arg Leu Phe Gly Pro Gly Lys Asn Val
260 265 270

Pro Ser Val Trp Arg Ser Ala Arg Arg Lys Arg Lys Lys Lys His Arg
275 280 285

Glu Leu Ile Gln Glu Glu Gln Ile Gln Glu Val Glu Cys Ser Val Glu
290 295 300

Ser Glu Val Ser Gln Lys Ser Leu Trp Asn Tyr Asp Tyr Ala Pro Pro
305 310 315 320

Pro Pro Pro Glu Gln Cys Leu Ser Asp Asp Glu Ile Thr Met Met Ala
325 330 335

Pro Val Glu Ser Lys Phe Ser Gln Ser Thr Gly Asp Ile Asp Lys Val
340 345 350

Thr Asp Thr Lys Pro Arg Val Ala Glu Trp Arg Tyr Gly Pro Ala Arg
355 360 365

Leu Trp Tyr Asp Met Leu Gly Val Pro Glu Asp Gly Ser Gly Phe Asp
370 375 380

Tyr Gly Phe Lys Leu Arg Lys Thr Glu His Glu Pro Val Ile Lys Ser
385 390 395 400

Arg Met Ile Glu Glu Phe Arg Lys Leu Glu Glu Asn Asn Gly Thr Asp
405 410 415

Leu Leu Ala Asp Glu Asn Phe Leu Met Val Thr Gln Leu His Trp Glu
420 425 430

Asp Asp Ile Ile Trp Asp Gly Glu Asp Val Lys His Lys Gly Thr Lys
435 440 445

Pro Gln Arg Ala Ser Leu Ala Gly Trp Leu Pro Ser Ser Met Thr Arg
450 455 460

Asn Ala Met Ala Tyr Asn Val Gln Gln Gly Phe Ala Ala Thr Leu Asp
465 470 475 480

Asp Asp Lys Pro Trp Tyr Ser Ile Phe Pro Ile Asp Asn Glu Asp Leu
485 490 495

Val Tyr Gly Arg Trp Glu Asp Asn Ile Ile Trp Asp Ala Gln Ala Met
500 505 510

Pro Arg Leu Leu Glu Pro Pro Val Leu Thr Leu Asp Pro Asn Asp Glu
515 520 525

Asn Leu Ile Leu Glu Ile Pro Asp Glu Lys Glu Glu Ala Thr Ser Asn
530 535 540

Ser Pro Ser Lys Glu Ser Lys Lys Glu Ser Ser Leu Lys Lys Ser Arg
545 550 555 560

Ile Leu Leu Gly Lys Thr Gly Val Ile Lys Glu Glu Pro Gln Gln Asn
565 570 575

Met Ser Gln Pro Glu Val Lys Asp Pro Trp Asn Leu Ser Asn Asp Glu
580 585 590

Tyr Tyr Tyr Pro Lys Gln Gln Gly Leu Arg Gly Thr Phe Gly Gly Asn
595 600 605

Ile Ile Gln His Ser Ile Pro Ala Val Glu Leu Arg Gln Pro Phe Phe
610 615 620

Pro Thr His Met Gly Pro Ile Lys Leu Arg Gln Phe His Arg Pro Pro
625 630 635 640

Leu Lys Lys Tyr Ser Phe Gly Ala Leu Ser Gln Pro Gly Pro His Ser
645 650 655

Val Gln Pro Leu Leu Lys His Ile Lys Lys Lys Ala Lys Met Arg Glu
660 665 670

Gln Glu Arg Gln Ala Ser Gly Gly Gly Glu Met Phe Phe Met Arg Thr
675 680 685

Pro Gln Asp Leu Thr Gly Lys Asp Gly Asp Leu Ile Leu Ala Glu Tyr
690 695 700

Ser Glu Glu Asn Gly Pro Leu Met Met Gln Val Gly Met Ala Thr Lys
705 710 715 720

Ile Lys Asn Tyr Tyr Lys Arg Lys Pro Gly Lys Asp Pro Gly Ala Pro
725 730 735

Asp Cys Lys Tyr Gly Glu Thr Val Tyr Cys His Thr Ser Pro Phe Leu
740 745 750

Gly Ser Leu His Pro Gly Gln Leu Leu Gln Ala Phe Glu Asn Asn Leu
755 760 765

Phe Arg Ala Pro Ile Tyr Leu His Lys Met Pro Glu Thr Asp Phe Leu
770 775 780

Ile Ile Arg Thr Arg Gln Gly Tyr Tyr Ile Arg Glu Leu Val Asp Ile
785 790 795 800

Phe Val Val Gly Gln Gln Cys Pro Leu Phe Glu Val Pro Gly Pro Asn
805 810 815

Ser Lys Arg Ala Asn Thr His Ile Arg Asp Phe Leu Gln Val Phe Ile
820 825 830

Tyr Arg Leu Phe Trp Lys Ser Lys Asp Arg Pro Arg Arg Ile Arg Met
835 840 845

Glu Asp Ile Lys Lys Ala Phe Pro Ser His Ser Glu Ser Ser Ile Arg
850 855 860

Lys Arg Leu Lys Leu Cys Ala Asp Phe Lys Arg Thr Gly Met Asp Ser
865 870 875 880

Asn Trp Trp Val Leu Lys Ser Asp Phe Arg Leu Pro Thr Glu Glu Glu
885 890 895

Ile Arg Ala Met Val Ser Pro Glu Gln Cys Cys Ala Tyr Tyr Ser Met
900 905 910

Ile Ala Ala Glu Gln Arg Leu Lys Asp Ala Gly Tyr Gly Glu Lys Ser
915 920 925

Phe Phe Ala Pro Glu Glu Glu Asn Glu Glu Asp Phe Gln Met Lys Ile
930 935 940

Asp Asp Glu Val Arg Thr Ala Pro Trp Asn Thr Thr Arg Ala Phe Ile
945 950 955 960

Ala Ala Met Lys Gly Lys Cys Leu Leu Glu Val Thr Gly Val Ala Asp
965 970 975

Pro Thr Gly Cys Gly Glu Gly Phe Ser Tyr Val Lys Ile Pro Asn Lys
980 985 990

Pro Thr Gln Gln Lys Asp Asp Lys Glu Pro Gln Pro Val Lys Lys Thr
995 1000 1005

Val Thr Gly Thr Asp Ala Asp Leu Arg Arg Leu Ser Leu Lys Asn Ala
1010 1015 1020

Lys Gln Leu Leu Arg Lys Phe Gly Val Pro Glu Glu Glu Ile Lys Lys
1025 1030 1035 1040

Leu Ser Arg Trp Glu Val Ile Asp Val Val Arg Thr Met Ser Thr Glu
1045 1050 1055

Gln Ala Arg Ser Gly Glu Gly Pro Met Ser Lys Phe Ala Arg Gly Ser
1060 1065 1070

Arg Phe Ser Val Ala Glu His Gln Glu Arg Tyr Lys Glu Glu Cys Gln
1075 1080 1085

Arg Ile Phe Asp Leu Gln Asn Lys Val Leu Ser Ser Thr Glu Val Leu
1090 1095 1100

Ser Thr Asp Thr Asp Ser Ser Ser Ala Glu Asp Ser Asp Phe Glu Glu
1105 1110 1115 1120

Met Gly Lys Asn Ile Glu Asn Met Leu Gln Asn Lys Lys Thr Ser Ser
1125 1130 1135

Gln Leu Ser Arg Glu Arg Glu Glu Gln Glu Arg Lys Glu Leu Gln Arg
1140 1145 1150

Met Leu Leu Ala Ala Gly Ser Ala Ala Ser Gly Asn Asn His Arg Asp
1155 1160 1165

Asp Asp Thr Ala Ser Val Thr Ser Leu Asn Ser Ser Ala Thr Gly Arg
1170 1175 1180

Cys Leu Lys Ile Tyr Arg Thr Phe Arg Asp Glu Glu Gly Lys Glu Tyr
1185 1190 1195 1200

Val Arg Cys Glu Thr Val Arg Lys Pro Ala Val Ile Asp Ala Tyr Val
1205 1210 1215

Arg Ile Arg Thr Thr Lys Asp Glu Glu Phe Ile Arg Lys Phe Ala Leu
1220 1225 1230

Phe Asp Glu Gln His Arg Glu Glu Met Arg Lys Glu Arg Arg Arg Ile
1235 1240 1245

Gln Glu Gln Leu Arg Arg Leu Lys Arg Asn Gln Glu Lys Glu Lys Leu
1250 1255 1260

Lys Gly Pro Pro Glu Lys Lys Pro Lys Lys Met Lys Glu Arg Pro Asp
1265 1270 1275 1280

Leu Lys Leu Lys Cys Gly Ala Cys Gly Ala Ile Gly His Met Arg Thr
1285 1290 1295

Asn Lys Phe Cys Pro Leu Tyr Tyr Gln Thr Asn Ala Pro Pro Ser Asn
1300 1305 1310

Pro Val Ala Met Thr Glu Glu Gln Glu Glu Glu Leu Glu Lys Thr Val
1315 1320 1325

Ile His Asn Asp Asn Glu Glu Leu Ile Lys Val Glu Gly Thr Lys Ile
1330 1335 1340

Val Leu Gly Lys Gln Leu Ile Glu Ser Ala Asp Glu Val Arg Arg Lys
1345 1350 1355 1360

Ser Leu Val Leu Lys Phe Pro Lys Gln Gln Leu Pro Pro Lys Lys Lys
1365 1370 1375

Arg Arg Val Gly Thr Thr Val His Cys Asp Tyr Leu Asn Arg Pro His
1380 1385 1390

Lys Ser Ile His Arg Arg Arg Thr Asp Pro Met Val Thr Leu Ser Ser
1395 1400 1405

Ile Leu Glu Ser Ile Ile Asn Asp Met Arg Asp Leu Pro Asn Thr Tyr
1410 1415 1420

Pro Phe His Thr Pro Val Asn Ala Lys Val Val Lys Asp Tyr Tyr Lys
1425 1430 1435 1440

Ile Ile Thr Arg Pro Met Asp Leu Gln Thr Leu Arg Glu Asn Val Arg
1445 1450 1455

Lys Arg Leu Tyr Pro Ser Arg Glu Glu Phe Arg Glu His Leu Glu Leu
1460 1465 1470

Ile Val Lys Asn Ser Ala Thr Tyr Asn Gly Pro Lys His Ser Leu Thr
1475 1480 1485

Gln Ile Ser Gln Ser Met Leu Asp Leu Cys Asp Glu Lys Leu Lys Glu
1490 1495 1500

Lys Glu Asp Lys Leu Ala Arg Leu Glu Lys Ala Ile Asn Pro Leu Leu
1505 1510 1515 1520

Asp Asp Asp Asp Gln Val Ala Phe Ser Phe Ile Leu Asp Asn Ile Val
1525 1530 1535

Thr Gln Lys Met Met Ala Val Pro Asp Ser Trp Pro Phe His His Pro
1540 1545 1550

Val Asn Lys Lys Phe Val Pro Asp Tyr Tyr Lys Val Ile Val Asn Pro
1555 1560 1565

Met Asp Leu Glu Thr Ile Arg Lys Asn Ile Ser Lys His Lys Tyr Gln
1570 1575 1580

Ser Arg Glu Ser Phe Leu Asp Asp Val Asn Leu Ile Leu Ala Asn Ser
1585 1590 1595 1600

Val Lys Tyr Asn Gly Pro Glu Ser Gln Tyr Thr Lys Thr Ala Gln Glu
1605 1610 1615

Ile Val Asn Val Cys Tyr Gln Thr Leu Thr Glu Tyr Asp Glu His Leu
1620 1625 1630

Thr Gln Leu Glu Lys Asp Ile Cys Thr Ala Lys Glu Ala Ala Leu Glu
1635 1640 1645

Glu Ala Glu Leu Glu Ser Leu Asp Pro Met Thr Pro Gly Pro Tyr Thr
1650 1655 1660

Pro Gln Pro Pro Asp Leu Tyr Asp Thr Asn Thr Ser Leu Ser Met Ser
1665 1670 1675 1680

Arg Asp Ala Ser Val Phe Gln Asp Glu Ser Asn Met Ser Val Leu Asp
1685 1690 1695

Ile Pro Ser Ala Thr Pro Glu Lys Gln Val Thr Gln Glu Gly Glu Asp
1700 1705 1710

Gly Asp Gly Asp Leu Ala Asp Glu Glu Glu Gly Thr Val Gln Gln Pro
1715 1720 1725

Gln Ala Ser Val Leu Tyr Glu Asp Leu Leu Met Ser Glu Gly Glu Asp
1730 1735 1740

Asp Glu Glu Asp Ala Gly Ser Asp Glu Glu Gly Asp Asn Pro Phe Ser
1745 1750 1755 1760

Ala Ile Gln Leu Ser Glu Ser Gly Ser Asp Ser Asp Val Gly Ser Gly
1765 1770 1775

Gly Ile Arg Pro Lys Gln Pro Arg Met Leu Gln Glu Asn Thr Arg Met
1780 1785 1790

Asp Met Glu Asn Glu Glu Ser Met Met Ser Tyr Glu Gly Asp Gly Gly
1795 1800 1805

Glu Ala Ser His Gly Leu Glu Asp Ser Asn Ile Ser Tyr Gly Ser Tyr
1810 1815 1820

Glu Glu Pro Asp Pro Lys Ser Asn Thr Gln Asp Thr Ser Phe Ser Ser
1825 1830 1835 1840

Ile Gly Gly Tyr Glu Val Ser Glu Glu Glu Glu Asp Glu Glu Glu Glu
1845 1850 1855

Glu Gln Arg Ser Gly Pro Ser Val Leu Ser Gln Val His Leu Ser Glu
1860 1865 1870

Asp Glu Glu Asp Ser Glu Asp Phe His Ser Ile Ala Gly Asp Ser Asp
1875 1880 1885

Leu Asp Ser Asp Glu
1890

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3182 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 972..3002

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGAGTTTTTT TTTTTTTTTT TTTTACAAGA GCACAAATCC ACATTTATTT ATTGATTTTT
60

CGTTAGTTTA AATCCTTGAG GGGTACAGCA TCACTCGGAT TCTGTGTCCA ATGGCCTTAG
120

CAGGAAGATT GCTTCGGAAT TTGGCACGAA CCATGCCACT GTTTCCATGG GCCCGAGTTA
180

CTTTTCCCCA GATGACTCTG GTTTTGTTTG GTTTGCCGCC AGGAGTGACT GTGTTGTTCT
240

TTGCTTTATA TACATAAGCG CATCTCTTGC CCAAATAGAA TTCTGTTTCA TCTCGGGCGT
300

AAACACCTTC AATTTTAAGA AGAGCTGTGT GCTCCCTTTG GTTCCGGAGA CCCCGCTTAT
360

AGCCAGCAAA AATGGCCTTG GACCACAGCC TTCCAGACAT AGTTCCTTTT AGAAGTCCCG
420

TTCCCAGCAG GCCTCCACAG GAGCCAAGAT GGCGCCGAGC CGGGTGAGCA GCGTCTCGGC
480

TGCCGCTAGA GTTTTCCTGC TCCCCGCGCT CGGGTGGCGG GGGCGGGTCT GAGTGGTACC
540

CCGGAGGAGA CCCTTTGAAG GTCCCTTGTG GGGACTGGAA AGAGGACGGT TGGTTGTGTG
600

TCTGTGCTCG TGGGGACCCC GTGTGTGTGC CTGCATTGGA GAGATGTTGC AGGAGATGGG
660

GTGGGCTCTC TGAACCTCCT TTCGCGCTGC CCGGGGATCT TCGACCTGCT TCTCTGCTGG
720

GATCTCGCTT AAGTTAACCC TTCCCTGGGA CGCCTTCCTG CCGCCTCCAC TGATCTGAGG
780

AGATCCTGTG ACTGTAGCGT GTTTTATGAG CCTTTACTGG CAGAGGGTAC CGCCGGGTAT
840

TGAAGGATTC GTAGGAGTTC GCCAGGGAAG TGGGACACGA CCCCCTCTTG TAAACCCGGC
900

GCCAGGCACA GAGGTCTCGG TCTCTCCACC GGGGGCTTCA TCCTTCCAGG GAGGAGAAGA

960

GGGACTCCAG A ATG GCT GAG GAG AAG AAG CTG AAG CTT AGC AAC ACT GTG
1010
Met Ala Glu Glu Lys Lys Leu Lys Leu Ser Asn Thr Val
1 5 10

CTG CCC TCG GAG TCC ATG AAG GTG GTG GCT GAA TCC ATG GGC ATC GCC
1058
Leu Pro Ser Glu Ser Met Lys Val Val Ala Glu Ser Met Gly Ile Ala
15 20 25

CAG ATT CAG GAG GAG ACC TGC CAG CTG CTA ACG GAT GAG GTC AGC TAC
1106
Gln Ile Gln Glu Glu Thr Cys Gln Leu Leu Thr Asp Glu Val Ser Tyr
30 35 40
45

CGC ATC AAA GAG ATC GCA CAG GAT GCC TTG AAG TTC ATG CAC ATG GGG
1154
Arg Ile Lys Glu Ile Ala Gln Asp Ala Leu Lys Phe Met His Met Gly
50 55
60

AAG CGG CAG AAG CTC ACC ACC AGT GAC ATT GAC TAC GCC TTG AAG CTA
1202
Lys Arg Gln Lys Leu Thr Thr Ser Asp Ile Asp Tyr Ala Leu Lys Leu
65 70 75

AAG AAT GTC GAG CCA CTC TAT GGC TTC CAC GCC CAG GAC TTC ATT CCT
1250
Lys Asn Val Glu Pro Leu Tyr Gly Phe His Ala Gln Asp Phe Ile Pro
80 85 90

TTC CGC TTC GCC TCT GGT GGG GGC CGG GAG CTT TAC TTC TAT GAG GAG

1298
Phe Arg Phe Ala Ser Gly Gly Gly Arg Glu Leu Tyr Phe Tyr Glu Glu
95 100 105

AAG GAG GTT GAT CTG AGC GAC ATC ATC AAT ACC CCT CTG CCC CGG GTG
1346
Lys Glu Val Asp Leu Ser Asp Ile Ile Asn Thr Pro Leu Pro Arg Val
110 115 120
125

CCC CTG GAC GTC TGC CTC AAA GCT CAT TGG CTG AGC ATC GAG GGC TGC
1394
Pro Leu Asp Val Cys Leu Lys Ala His Trp Leu Ser Ile Glu Gly Cys
130 135 140

CAG CCA GCT ATC CCC GAG AAC CCG CCC CCA GCT CCC AAA GAG CAA CAG
1442
Gln Pro Ala Ile Pro Glu Asn Pro Pro Pro Ala Pro Lys Glu Gln Gln
145 150 155

AAG GCT GAA GCC ACA GAA CCC CTG AAG TCA GCC AAG CCA GGC CAG GAG
1490
Lys Ala Glu Ala Thr Glu Pro Leu Lys Ser Ala Lys Pro Gly Gln Glu
160 165 170

GAA GAC GGA CCC CTG AAG GGC AAA GGT CAA GGG GCC ACC ACA GCC GAC
1538
Glu Asp Gly Pro Leu Lys Gly Lys Gly Gln Gly Ala Thr Thr Ala Asp
175 180 185

GGC AAA GGG AAA GAG AAG AAG GCG CCG CCC TTG CTG GAG GGG GCC CCC
1586
Gly Lys Gly Lys Glu Lys Lys Ala Pro Pro Leu Leu Glu Gly Ala Pro
190 195 200
205

```
TTG CGA CTG AAG CCC CGG AGC ATC CAC GAG TTG TCT GTG GAG CAG
CAG
  1634
Leu Arg Leu Lys Pro Arg Ser Ile His Glu Leu Ser Val Glu Gln
Gln
                210                 215                 220


CTC TAC TAC AAG GAG ATC ACC GAG GCC TGC GTG GGC TCC TGC GAG
GCC
  1682
Leu Tyr Tyr Lys Glu Ile Thr Glu Ala Cys Val Gly Ser Cys Glu
Ala
            225                 230                 235


AAG AGG GCG GAA GCC CTG CAA AGC ATT GCC ACG GAC CCT GGA CTG
TAT
  1730
Lys Arg Ala Glu Ala Leu Gln Ser Ile Ala Thr Asp Pro Gly Leu
Tyr
        240                 245                 250


CAG ATG CTG CCA CGG TTC AGT ACC TTT ATC TCG GAG GGG GTC CGT
GTG
  1778
Gln Met Leu Pro Arg Phe Ser Thr Phe Ile Ser Glu Gly Val Arg
Val
    255                 260                 265


AAC GTG GTT CAG AAC AAC CTG GCC CTA CTC ATC TAC CTG ATG CGT
ATG
  1826
Asn Val Val Gln Asn Asn Leu Ala Leu Leu Ile Tyr Leu Met Arg
Met
270                 275                 280
  285

GTG AAA GCG CTG ATG GAC AAC CCC ACG CTC TAT CTA GAA AAA TAC
GTC
  1874
Val Lys Ala Leu Met Asp Asn Pro Thr Leu Tyr Leu Glu Lys Tyr
Val
                290                 295                 300


CAT GAG CTG ATT CCA GCT GTG ATG ACC TGC ATC GTG AGC AGA CAG
TTG
  1922
His Glu Leu Ile Pro Ala Val Met Thr Cys Ile Val Ser Arg Gln
Leu
```

305 310 315

TGC CTG CGA CCA GAT GTG GAC AAT CAC TGG GCA CTC CGA GAC TTT GCT
1970
Cys Leu Arg Pro Asp Val Asp Asn His Trp Ala Leu Arg Asp Phe Ala
320 325 330

GCC CGC CTG GTG GCC CAG ATC TGC AAG CAT TTT AGC ACA ACC ACT AAC
2018
Ala Arg Leu Val Ala Gln Ile Cys Lys His Phe Ser Thr Thr Thr Asn
335 340 345

AAC ATC CAG TCC CGG ATC ACC AAG ACC TTC ACC AAG AGC TGG GTG GAC
2066
Asn Ile Gln Ser Arg Ile Thr Lys Thr Phe Thr Lys Ser Trp Val Asp
350 355 360
365

GAG AAG ACG CCC TGG ACG ACT CGT TAT GGC TCC ATC GCA GGC TTG GCT
2114
Glu Lys Thr Pro Trp Thr Thr Arg Tyr Gly Ser Ile Ala Gly Leu Ala
370 375 380

GAG CTG GGA CAC GAT GTT ATC AAG ACT CTG ATT CTG CCC CGG CTG CAG
2162
Glu Leu Gly His Asp Val Ile Lys Thr Leu Ile Leu Pro Arg Leu Gln
385 390 395

ACC TTC ACC AAG AGC TGG GTG GAC GAG AAG ACG CCC TGG ACG ACT CGT
2210
Thr Phe Thr Lys Ser Trp Val Asp Glu Lys Thr Pro Trp Thr Thr Arg
400 405 410

TAT GGC TCC AGG ATT GGA GCA GAC CAT GTG CAG AGC CTC CTG CTG AAA
2258

```
Tyr Gly Ser Arg Ile Gly Ala Asp His Val Gln Ser Leu Leu Leu
Lys
    415                 420                 425


CAC TGT GCT CCT GTT CTG GCA AAG CTG CGC CCA CCG CCT GAC AAT
CAG
  2306
His Cys Ala Pro Val Leu Ala Lys Leu Arg Pro Pro Pro Asp Asn
Gln
430                 435                 440
  445

GAC GCC TAT CGG GCA GAA TTC GGG TCC CTT GGG CCC CTC CTC TGC
TCC
  2354
Asp Ala Tyr Arg Ala Glu Phe Gly Ser Leu Gly Pro Leu Leu Cys
Ser
                450                 455                 460


CAG GTG GTC AAG GCT CGG GCC CAG GCT GCT CTG CAG GCT CAG CAG
GTC
  2402
Gln Val Val Lys Ala Arg Ala Gln Ala Ala Leu Gln Ala Gln Gln
Val
            465                 470                 475


AAC AGG ACC ACT CTG ACC ATC ACG CAG CCC CGG CCC ACG CTG ACC
CTC
  2450
Asn Arg Thr Thr Leu Thr Ile Thr Gln Pro Arg Pro Thr Leu Thr
Leu
        480                 485                 490


TCG CAG GCC CCA CAG CCT GGC CCT CGC ACC CCT GGC TTG CTG AAG
GTT
  2498
Ser Gln Ala Pro Gln Pro Gly Pro Arg Thr Pro Gly Leu Leu Lys
Val
    495                 500                 505


CCT GGC TCC ATC GCA CTT CCT GTC CAG ACA CTG GTG TCT GCA CGA
GCG
  2546
Pro Gly Ser Ile Ala Leu Pro Val Gln Thr Leu Val Ser Ala Arg
Ala
510                 515                 520
  525
```

```
GCT GCC CCA CCA CAG CCT TCC CCT CCT CCC ACC AAG TTT ATT GTA
ATG
  2594
Ala Ala Pro Pro Gln Pro Ser Pro Pro Pro Thr Lys Phe Ile Val
Met
                530                 535                 540


TCA TCG TCC TCC AGC GCC CCA TCC ACC CAG CAG GTC CTG TCC CTC
AGC
  2642
Ser Ser Ser Ser Ser Ala Pro Ser Thr Gln Gln Val Leu Ser Leu
Ser
            545                 550                 555


ACC TCG GCC CCC GGC TCA GGT TCC ACC ACC ACT TCG CCC GTC ACC
ACC
  2690
Thr Ser Ala Pro Gly Ser Gly Ser Thr Thr Thr Ser Pro Val Thr
Thr
        560                 565                 570


ACC GTC CCC AGC GTG CAG CCC ATC GTC AAG TTG GTC TCC ACC GCC
ACC
  2738
Thr Val Pro Ser Val Gln Pro Ile Val Lys Leu Val Ser Thr Ala
Thr
    575                 580                 585


ACC GCA CCC CCC AGC ACT GCT CCC TCT GGT CCT GGG AGT GTC CAG
AAG
  2786
Thr Ala Pro Pro Ser Thr Ala Pro Ser Gly Pro Gly Ser Val Gln
Lys
590                 595                 600
  605

TAC ATC GTG GTC TCA CTT CCC CCA ACA GGG GAG GGC AAA GGA GGC
CCC
  2834
Tyr Ile Val Val Ser Leu Pro Pro Thr Gly Glu Gly Lys Gly Gly
Pro
                610                 615                 620


ACC TCC CAT CCT TCT CCA GTT CCT CCC CCG GCA TCG TCC CCG TCC
CCA
  2882
Thr Ser His Pro Ser Pro Val Pro Pro Pro Ala Ser Ser Pro Ser
Pro
```

625 630 635

CTC AGC GGC AGT CGG GTT TGT GGG GGG AAG CAG GAG GCT GGG GAC AGT
2930
Leu Ser Gly Ser Arg Val Cys Gly Gly Lys Gln Glu Ala Gly Asp Ser
640 645 650

CCC CCT CCA GCT CCA GGG ACT CCA AAA GCC AAT GGC TCC CAG CCC AAC
2978
Pro Pro Pro Ala Pro Gly Thr Pro Lys Ala Asn Gly Ser Gln Pro Asn
655 660 665

TGC GGC TCC CCT CAG CCT GCT CCG TGATGCTCCA CCTGCCAGCC CCCGGATTCC
3032
Cys Gly Ser Pro Gln Pro Ala Pro
670 675

CACACATGCA GACATGTACA CACGTGCACG TACACACATG CATGCTCGCT AAGCGGAAGG
3092

AAGTTGTAGA TTGCTTCCTT CATGTCACTT TCTTTTTAGA TATTGTACAG CCAGTTTCTC
3152

AGAATAAAAG TTTGGTTTGT AAAAAAAAAA

3182

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 677 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Ala Glu Glu Lys Lys Leu Lys Leu Ser Asn Thr Val Leu Pro Ser

1 5 10 15

Glu Ser Met Lys Val Val Ala Glu Ser Met Gly Ile Ala Gln Ile Gln
20 25 30

Glu Glu Thr Cys Gln Leu Leu Thr Asp Glu Val Ser Tyr Arg Ile Lys
35 40 45

Glu Ile Ala Gln Asp Ala Leu Lys Phe Met His Met Gly Lys Arg Gln
50 55 60

Lys Leu Thr Thr Ser Asp Ile Asp Tyr Ala Leu Lys Leu Lys Asn Val
65 70 75 80

Glu Pro Leu Tyr Gly Phe His Ala Gln Asp Phe Ile Pro Phe Arg Phe
85 90 95

Ala Ser Gly Gly Gly Arg Glu Leu Tyr Phe Tyr Glu Glu Lys Glu Val
100 105 110

Asp Leu Ser Asp Ile Ile Asn Thr Pro Leu Pro Arg Val Pro Leu Asp
115 120 125

Val Cys Leu Lys Ala His Trp Leu Ser Ile Glu Gly Cys Gln Pro Ala
130 135 140

Ile Pro Glu Asn Pro Pro Pro Ala Pro Lys Glu Gln Gln Lys Ala Glu
145 150 155 160

Ala Thr Glu Pro Leu Lys Ser Ala Lys Pro Gly Gln Glu Glu Asp Gly
165 170 175

Pro Leu Lys Gly Lys Gly Gln Gly Ala Thr Thr Ala Asp Gly Lys Gly
180 185 190

Lys Glu Lys Lys Ala Pro Pro Leu Leu Glu Gly Ala Pro Leu Arg Leu
195 200 205

Lys Pro Arg Ser Ile His Glu Leu Ser Val Glu Gln Gln Leu Tyr Tyr
210 215 220

Lys Glu Ile Thr Glu Ala Cys Val Gly Ser Cys Glu Ala Lys Arg Ala
225 230 235 240

Glu Ala Leu Gln Ser Ile Ala Thr Asp Pro Gly Leu Tyr Gln Met Leu
245 250 255

Pro Arg Phe Ser Thr Phe Ile Ser Glu Gly Val Arg Val Asn Val Val
260 265 270

Gln Asn Asn Leu Ala Leu Leu Ile Tyr Leu Met Arg Met Val Lys Ala
275 280 285

Leu Met Asp Asn Pro Thr Leu Tyr Leu Glu Lys Tyr Val His Glu Leu
290 295 300

Ile Pro Ala Val Met Thr Cys Ile Val Ser Arg Gln Leu Cys Leu Arg
305 310 315 320

Pro Asp Val Asp Asn His Trp Ala Leu Arg Asp Phe Ala Ala Arg Leu
325 330 335

Val Ala Gln Ile Cys Lys His Phe Ser Thr Thr Thr Asn Asn Ile Gln
340 345 350

Ser Arg Ile Thr Lys Thr Phe Thr Lys Ser Trp Val Asp Glu Lys Thr
355 360 365

Pro Trp Thr Thr Arg Tyr Gly Ser Ile Ala Gly Leu Ala Glu Leu Gly
370 375 380

His Asp Val Ile Lys Thr Leu Ile Leu Pro Arg Leu Gln Thr Phe Thr
385 390 395 400

Lys Ser Trp Val Asp Glu Lys Thr Pro Trp Thr Thr Arg Tyr Gly Ser
405 410 415

Arg Ile Gly Ala Asp His Val Gln Ser Leu Leu Leu Lys His Cys Ala
420 425 430

Pro Val Leu Ala Lys Leu Arg Pro Pro Pro Asp Asn Gln Asp Ala Tyr
435 440 445

Arg Ala Glu Phe Gly Ser Leu Gly Pro Leu Leu Cys Ser Gln Val Val
450 455 460

Lys Ala Arg Ala Gln Ala Ala Leu Gln Ala Gln Gln Val Asn Arg Thr
465 470 475 480

Thr Leu Thr Ile Thr Gln Pro Arg Pro Thr Leu Thr Leu Ser Gln Ala
485 490 495

Pro Gln Pro Gly Pro Arg Thr Pro Gly Leu Leu Lys Val Pro Gly Ser
500 505 510

Ile Ala Leu Pro Val Gln Thr Leu Val Ser Ala Arg Ala Ala Ala Pro
515 520 525

Pro Gln Pro Ser Pro Pro Pro Thr Lys Phe Ile Val Met Ser Ser Ser
530 535 540

Ser Ser Ala Pro Ser Thr Gln Gln Val Leu Ser Leu Ser Thr Ser Ala
545 550 555 560

Pro Gly Ser Gly Ser Thr Thr Thr Ser Pro Val Thr Thr Thr Val Pro
565 570 575

Ser Val Gln Pro Ile Val Lys Leu Val Ser Thr Ala Thr Thr Ala Pro
580 585 590

Pro Ser Thr Ala Pro Ser Gly Pro Gly Ser Val Gln Lys Tyr Ile Val
595 600 605

Val Ser Leu Pro Pro Thr Gly Glu Gly Lys Gly Gly Pro Thr Ser His
610 615 620

Pro Ser Pro Val Pro Pro Pro Ala Ser Ser Pro Ser Pro Leu Ser Gly
625 630 635 640

Ser Arg Val Cys Gly Gly Lys Gln Glu Ala Gly Asp Ser Pro Pro Pro
645 650 655

Ala Pro Gly Thr Pro Lys Ala Asn Gly Ser Gln Pro Asn Cys Gly Ser
660 665 670

Pro Gln Pro Ala Pro
675

dTAFII250 amino acid sequence

MGPGCDLLLRTAATITAAAIMSDTDSDEDSAGGGPFSLAGFLFGNINGAGQLEGESV
LDDECKKHLAGLGALGLGSLITELTANEELTGTDGALVNDEGWVRSTEDAVDYSDIN
EVAEDESRRYQQTMGSLQPLCHSDYDEDDYDADCEDIDCKLMPPPPPPPGPMKKDKD
QDSITGEKVDFSSSSDSESEMGPQEATQAESEDGKLTLPLAGIMQHDATKLLPSVTEL
FPEFRPGKVLRFLRLFGPGKNVPSVWRSARRKRKKKHRELIQEEQIQEVECSVESEVS
QKSLWNYDYAPPPPPEQCLSDDEITMMAPVESKFSQSTGDIDKVTDTKPRVAEWRY
GPARLWYDMLGVPEDGSGFDYGFKLRKTEHEPVIKSRMIEEFRKLEENNGTDLLADE
NFLMVTQLHWEDDIIWDGEDVKHKGTKPQRASLAGWLPSSMTRNAMAYNVQQGF
AATLDDDKPWYSIFPIDNEDLVYGRWEDNIIWDAQAMPRLLEPPVLTLDPNDENLI
LEIPDEKEEATSNSPSKESKKESSLKKSRILLGKTGVIKEEPQQNMSQPEVKDPWNLSN
DEYYYPKQQGLRGTFGGNIIQHSIPAVELRQPFFPTHMGPIKLRQFHRPPLKKYSFGA
LSQPGPHSVQPLLKHIKKKAKMREQERQASGGGEMFFMRTPQDLTGKDGDLILAEYS
EENGPLMMQVGMATKIKNYYKRKPGKDPGAPDCKYGETVYCHTSPFLGSLHPGQLL
QAFENNLFRAPIYLHKMPETDFLIIRTRQGYYIRELVDIFVVGQQCPLFEVPGPNSKR
ANTHIRDFLQVFIYRLFWKSKDRPRRIRMEDIKKAFPSHSESSIRKRLKLCADFKRTG
MDSNWWVLKSDFRLPTEEEIRAMVSPEQCCAYYSMIAAEQRLKDAGYGEKSFFAPE
EENEEDFQMKIDDEVRTAPWNTTRAFIAAMKGKCLLEVTGVADPTGCGEGFSYVKI
PNKPTQQKDDKEPQPVKKTVTGTDADLRRLSLKNAKQLLRKFGVPEEEIKKLSRWEV
IDVVRTMSTEQARSGEGPMSKFARGSRFSVAEHQERYKEECQRIFDLQNKVLSSTEVL
STDTDSSSAEDSDFEEMGKNIENMLQNKKTSSQLSREREEQERKELQRMLLAAGSAAS
GNNHRDDDTASVTSLNSSATGRCLKIYRTFRDEEGKEYVRCETVRKPAVIDAYVRIR
TTKDEEFIRKFALFDEQHREEMRKERRRIQEQLRRLKRNQEKEKLKGPPEKKPKKMKER
PDLKLKCGACGAIGHMRTNKFCPLYYQTNAPPSNPVAMTEEQEEELEKTVIHNDNEE
LIKVEGTKIVLGKQLIESADEVRRKSLVLKFPKQQLPPKKKRRVGTTVHCDYLNRPHK
SIHRRRTDPMVTLSSILESIINDMRDLPNTYPFHTPVNAKVVKDYYKIITRPMDLQT
LRENVRKRLYPSREEFREHLELIVKNSATYNGPKHSLTQISQSMLDLCDEKLKEKEDKL
ARLEKAINPLLDDDDQVAFSFILDNIVTQKMMAVPDSWPFHHPVNKKFVPDYYKV
IVNPMDLETIRKNISKHKYQSRESFLDDVNLILANSVKYNGPESQYTKTAQEIVNVCY
QTLTEYDEHLTQLEKDICTAKEAALEEAELESLDPMTPGPYTPQPPDLYDTNTSLSMS
RDASVFQDESNMSVLDIPSATPEKQVTQEGEDGDGDLADEEEGTVQQPQASVLYEDL
LMSEGEDDEEDAGSDEEGDNPFSAIQLSESGSDSDVGSGGIRPKQPRMLQENTRMDME
NEESMMSYEGDGGEASHGLEDSNISYGSYEEPDPKSNTQDTSFSSIGGYEVSEEEEDEEE
EEQRSGPSVLSQVHLSEDEEDSEDFHSIAGDSDLDSDE

Sequence Range: 1 to 2214

```
          5     10    15     20     25     30     35     40    45
                *           *           *           *
AGA GGT GGT GCA GGC GGC GCC CCC GGC GGC GCA GAC CCT GGC GCC AGC
Arg Gly Gly Ala Gly Gly Ala Pro Gly Gly Ala Asp Pro Gly Ala Ser>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 50     55    60     65     70    75     80     85    90     95
  *           *           *           *           *
GGC CCG GCC AGC ACG GCG GCC AGC ATG GTC ATC GGG CCA ACT ATG CAA
Gly Pro Ala Ser Thr Ala Ala Ser Met Val Ile Gly Pro Thr Met Gln>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

   100   105    110    115   120    125    130   135    140
     *           *           *           *           *
GGG CGC TGC CCA GCC CGG CCG CCG TCC CGC CGC CCG CCC CCG GGA CCC
Gly Arg Cys Pro Ala Arg Pro Pro Ser Arg Arg Pro Pro Pro Gly Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

145   150    155    160   165    170    175   180    185    190
        *           *           *           *           *
CCA CCG GGC TGC CCA AAA GGC GCG GCC GGC GCA GTG ACC CAG AGC CTG
Pro Pro Gly Cys Pro Lys Gly Ala Ala Gly Ala Val Thr Gln Ser Leu>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 195    200    205   210    215    220   225    230    235   240
          *           *           *           *           *
TCC CGG ACG CCC ACG GCC ACC ACC AGC GGG ATT CGG GCC ACC CTG ACG
Ser Arg Thr Pro Thr Ala Thr Thr Ser Gly Ile Arg Ala Thr Leu Thr>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

   245    250   255    260    265   270    275    280   285
             *           *           *           *
CCC ACC GTG CTG GCC CCC CGC TTG CCG CAG CCG CCT CAG AAC CCG ACC
Pro Thr Val Leu Ala Pro Arg Leu Pro Gln Pro Pro Gln Asn Pro Thr>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

290    295   300    305   310    315   320    325   330    335
  *           *           *           *           *
AAC ATC CAG AAC TTC CAG CTG CCC CCA GGA ATG GTC CTC GTC CGA AGT
Asn Ile Gln Asn Phe Gln Leu Pro Pro Gly Met Val Leu Val Arg Ser>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

  340   345    350    355   360    365    370   375    380
     *           *           *           *           *
GAG AAT GGG CAG TTG TTA ATG ATT CCT CAG CAG GCC TTG GCC CAG ATG
Glu Asn Gly Gln Leu Leu Met Ile Pro Gln Gln Ala Leu Ala Gln Met>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

385   390    395    400   405    410    415   420    425    430
        *           *           *           *           *
CAG GCG CAG GCC CAT GCC CAG CCT CAG ACC ACC ATG GCG CCT CGC CCT
Gln Ala Gln Ala His Ala Gln Pro Gln Thr Thr Met Ala Pro Arg Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 435    440   445    450   455    460   465    470   475    480
          *           *           *           *           *
GCC ACC CCC ACA AGT GCC CCT CCC GTC CAG ATC TCC ACC GTA CAG GCA
Ala Thr Pro Thr Ser Ala Pro Pro Val Gln Ile Ser Thr Val Gln Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>
```

```
    485     490     495     500     505     510     515     520     525
                *       *       *       *       *       *
CCT GGA ACA CCT ATC ATT GCA CGG CAG GTG ACC CCA ACT ACC ATA ATT
Pro Gly Thr Pro Ile Ile Ala Arg Gln Val Thr Pro Thr Thr Ile Ile>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

530     535     540     545     550     555     560     565     570     575
  *               *               *               *               *
AAG CAA GTG TCT CAG GCC CAG ACA ACG GTG CAG CCC AGT GCA ACC CTG
Lys Gln Val Ser Gln Ala Gln Thr Thr Val Gln Pro Ser Ala Thr Leu>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

  580   585     590     595   600     605     610   615     620
    *               *               *               *
CAG CGC TCG CCC GGC GTC CAG CCT CAG CTC GTT CTG GGT GGC GCT GCC
Gln Arg Ser Pro Gly Val Gln Pro Gln Leu Val Leu Gly Gly Ala Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

625   630     635     640   645     650     655   660     665     670
        *               *               *               *               *
CAG ACG GCT TCA CTT GGG ACG GCG ACG GCT GTT CAG ACG GGG ACT CCT
Gln Thr Ala Ser Leu Gly Thr Ala Thr Ala Val Gln Thr Gly Thr Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

675     680     685     690     695     700     705     710     715     720
        *               *               *               *               *
CAG CGC ACG GTA CCA GGG GCG ACC ACC ACT TCC TCA GCT GCC ACG GAA
Gln Arg Thr Val Pro Gly Ala Thr Thr Thr Ser Ser Ala Ala Thr Glu>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

    725     730     735     740     745     750     755     760     765
            *               *               *               *
ACT ATG GAA AAC GTG AAG AAA TGT AAA AAT TTC CTA TCT ACG TTA ATA
Thr Met Glu Asn Val Lys Lys Cys Lys Asn Phe Leu Ser Thr Leu Ile>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

770     775     780     785     790     795     800     805     810     815
  *               *               *               *               *
AAA CTG GCT TCA TCT GGC AAG CAG TCT ACA GAG ACA GCA GCT AAT GTG
Lys Leu Ala Ser Ser Gly Lys Gln Ser Thr Glu Thr Ala Ala Asn Val>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

  820   825     830     835   840     845     850   855     860
    *               *               *               *
AAA GAG CTC GTG CAG AAT TTA CTG GAT GGA AAA ATA GAA GCA GAA GAT
Lys Glu Leu Val Gln Asn Leu Leu Asp Gly Lys Ile Glu Ala Glu Asp>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

865   870     875     880   885     890     895   900     905     910
        *               *               *               *               *
TTC ACA AGC AGG TTA TAC CGA GAA CTT AAT TCT TCA CCT CAA CCT TAC
Phe Thr Ser Arg Leu Tyr Arg Glu Leu Asn Ser Ser Pro Gln Pro Tyr>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

915     920     925     930     935     940     945     950     955     960
        *               *               *               *               *
CTT GTG CCT TTC CTG AAG AGG AGC TTA CCC GCC TTG AGA CAG CTG ACC
Leu Val Pro Phe Leu Lys Arg Ser Leu Pro Ala Leu Arg Gln Leu Thr>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

    965     970     975     980     985     990     995    1000    1005
            *               *               *               *
```

```
CCC GAC TCC GCG GCC TTC ATC CAG CAG AGC CAG CAG CAG CCG CCA CCG
Pro Asp Ser Ala Ala Phe Ile Gln Gln Ser Gln Gln Gln Pro Pro Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1010  1015  1020  1025  1030  1035  1040  1045  1050  1055
        *     *     *     *     *     *     *

CCC ACC TCG CAG GCC ACC ACT GCG CTC ACG GCC GTG GTG CTG AGT AGC
Pro Thr Ser Gln Ala Thr Thr Ala Leu Thr Ala Val Val Leu Ser Ser>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 1060  1065  1070  1075  1080  1085  1090  1095  1100
     *     *     *     *     *     *     *

TCG GTC CAG CGC ACG GCC GGG AAG ACG GCG GCC ACC GTG ACC AGT GCC
Ser Val Gln Arg Thr Ala Gly Lys Thr Ala Ala Thr Val Thr Ser Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1105  1110  1115  1120  1125  1130  1135  1140  1145  1150
        *     *     *     *     *     *     *

CTC CAG CCC CCT GTG CTC AGC CTC ACG CAG CCC ACG CAG GTC GGC GTC
Leu Gln Pro Pro Val Leu Ser Leu Thr Gln Pro Thr Gln Val Gly Val>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1155  1160  1165  1170  1175  1180  1185  1190  1195  1200
        *     *     *     *     *     *     *

GGC AAG CAG GGG CAA CCC ACA CCG CTG GTC ATC CAG CAG CCT CCG AAG
Gly Lys Gln Gly Gln Pro Thr Pro Leu Val Ile Gln Gln Pro Pro Lys>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 1205  1210  1215  1220  1225  1230  1235  1240  1245
     *     *     *     *     *

CCA GGA GCC CTG ATC CGG CCC CCG CAG GTG ACG TTG ACG CAG ACA CCC
Pro Gly Ala Leu Ile Arg Pro Pro Gln Val Thr Leu Thr Gln Thr Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1250  1255  1260  1265  1270  1275  1280  1285  1290  1295
        *     *     *     *     *     *

ATG GTC GCC CTG CGG CAG CCT CAC AAC CGG ATC ATG CTC ACC ACG CCT
Met Val Ala Leu Arg Gln Pro His Asn Arg Ile Met Leu Thr Thr Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 1300  1305  1310  1315  1320  1325  1330  1335  1340
     *     *     *     *     *

CAG CAG ATC CAG CTG AAC CCA CTG CAG CCA GTC CCT GTG GTG AAA CCC
Gln Gln Ile Gln Leu Asn Pro Leu Gln Pro Val Pro Val Val Lys Pro>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1345  1350  1355  1360  1365  1370  1375  1380  1385  1390
        *     *     *     *     *     *

GCC GTG TTA CCT GGA ACC AAA GCC CTT TCT GCT GTC TCG GCA CAA GCA
Ala Val Leu Pro Gly Thr Lys Ala Leu Ser Ala Val Ser Ala Gln Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

1395  1400  1405  1410  1415  1420  1425  1430  1435  1440
        *     *     *     *     *     *     *

GCT GCT GCA CAG AAA AAT AAA CTC AAG GAG CCT GGG GGA GGT TCG TTT
Ala Ala Ala Gln Lys Asn Lys Leu Lys Glu Pro Gly Gly Gly Ser Phe>
___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 1445  1450  1455  1460  1465  1470  1475  1480  1485
     *     *     *     *     *

CGG GAC GAT GAT GAC ATT AAT GAT GTT GCA TCG ATG GCT GGA GTA AAC
Arg Asp Asp Asp Asp Ile Asn Asp Val Ala Ser Met Ala Gly Val Asn>
```

```
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1490   1495  1500   1505   1510  1515   1520   1525  1530   1535
  *      *     *      *      *     *      *      *     *      *
TTG TCA GAA GAA AGT GCA AGA ATA TTA GCC ACG AAC TCT GAA TTG GTG
Leu Ser Glu Glu Ser Ala Arg Ile Leu Ala Thr Asn Ser Glu Leu Val>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

 1540  1545   1550   1555  1560   1565   1570  1575   1580
    *     *      *      *     *      *      *     *      *
GGC ACG CTA ACG CGG TCC TGT AAA GAT GAA ACC TTC CTC CTC CAA GCG
Gly Thr Leu Thr Arg Ser Cys Lys Asp Glu Thr Phe Leu Leu Gln Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1585  1590   1595   1600  1605   1610   1615  1620   1625   1630
       *      *      *     *      *      *     *      *      *
CCT TTG CAG AGA AGA ATA TTA GAA ATA GGT AAA AAA CAT GGT ATA ACG
Pro Leu Gln Arg Arg Ile Leu Glu Ile Gly Lys Lys His Gly Ile Thr>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1635   1640   1645  1650   1655   1660  1665   1670   1675  1680
       *      *     *      *      *     *      *      *     *
GAA TTA CAT CCA GAT GTA GTA AGT TAT GTA TCA CAT GCC ACG CAA CAA
Glu Leu His Pro Asp Val Val Ser Tyr Val Ser His Ala Thr Gln Gln>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

  1685   1690  1695   1700   1705  1710   1715   1720  1725
         *     *      *      *     *      *      *     *
AGG CTA CAG AAT CTT GTA GAG AAA ATA TCA GAA ACA GCT CAG CAG AAG
Arg Leu Gln Asn Leu Val Glu Lys Ile Ser Glu Thr Ala Gln Gln Lys>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1730   1735  1740   1745   1750  1755   1760   1765  1770   1775
  *      *     *      *      *     *      *      *     *      *
AAC TTT TCT TAC AAG GAT GAC GAC AGA TAT GAG CAG GCG AGT GAC GTC
Asn Phe Ser Tyr Lys Asp Asp Asp Arg Tyr Glu Gln Ala Ser Asp Val>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

 1780  1785   1790   1795  1800   1805   1810  1815   1820
    *     *      *      *     *      *      *     *      *
CGG GCA CAG CTC AAG TTT TTT GAA CAG CTT GAT CAA ATC GAA AAG CAG
Arg Ala Gln Leu Lys Phe Phe Glu Gln Leu Asp Gln Ile Glu Lys Gln>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1825  1830   1835   1840  1845   1850   1855  1860   1865   1870
       *      *      *     *      *      *     *      *      *
AGG AAG GAT GAG CAG GAG CGG GAG ATC CTG ATG AGG GCA GCA AAG TCT
Arg Lys Asp Glu Gln Glu Arg Glu Ile Leu Met Arg Ala Ala Lys Ser>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

1875   1880   1885  1890   1895   1900  1905   1910   1915  1920
       *      *     *      *      *     *      *      *     *
CGG TCA AGA CAA GAA GAT CCA GAA CAG TTA AGG CTG AAA CAG AAG GCA
Arg Ser Arg Gln Glu Asp Pro Glu Gln Leu Arg Leu Lys Gln Lys Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>

  1925   1930  1935   1940   1945  1950   1955   1960  1965
         *     *      *      *     *      *      *     *
AAG GAG ATG CAG CAA CAG GAA CTG GCA CAA ATG AGA CAG CGG GAC GCC
Lys Glu Met Gln Gln Gln Glu Leu Ala Gln Met Arg Gln Arg Asp Ala>
___a___a___a____TRANSLATION OF HTAF130 DNA (A)_a___a___a___a___>
```

```
1970   1975  1980   1985   1990  1995   2000   2005  2010   2015
  *            *             *            *             *
 AAC CTC ACA GCA CTA GCA GCG ATC GGG CCC AGG AAA AAG AGG AAA GTG
 Asn Leu Thr Ala Leu Ala Ala Ile Gly Pro Arg Lys Lys Arg Lys Val>
 ___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

  2020  2025   2030   2035  2040   2045   2050  2055   2060
    *            *             *            *             *
 GAC TGT CCG GGG CCG GGC TCA GGA GCA GAG GGG TCG GGC CCC GGC TCA
 Asp Cys Pro Gly Pro Gly Ser Gly Ala Glu Gly Ser Gly Pro Gly Ser>
 ___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

2065  2070   2075   2080  2085   2090  2095  2100   2105   2110
      *            *             *            *             *
 GTG GTC CCA GGC AGC TCG GGT GTC GGA ACC CCC AGA CAG TTC ACG CGA
 Val Val Pro Gly Ser Ser Gly Val Gly Thr Pro Arg Gln Phe Thr Arg>
 ___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

 2115   2120   2125  2130   2135   2140  2145   2150   2155  2160
        *            *             *            *             *
 CAA AGA ATC ACG CGG GTC AAC CTC AGG GAC CTC ATA TTT TGT TTA GAA
 Gln Arg Ile Thr Arg Val Asn Leu Arg Asp Leu Ile Phe Cys Leu Glu>
 ___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

   2165   2170  2175   2180   2185  2190   2195   2200  2205
          *            *             *            *
 AAT GAA CGT GAG ACA AGC CAT TCA CTG CTG CTC TAC AAA GCA TTC CTT
 Asn Glu Arg Glu Thr Ser His Ser Leu Leu Leu Tyr Lys Ala Phe Leu>
 ___a___a___a____TRANSLATION OF HTAF130 DNA [A]_a___a___a___a___>

2210
  *
 AAG TGA
 Lys ***>
 ___a___>
```

```
    5   10   15   20   25   30   35   40   45   50
    *    *    *    *    *    *    *    *    *    *
RGGAGGAPGG ADPGASGPAS TAASMVIGPT MQGRCPARPP SRRPPPGPPP

   55   60   65   70   75   80   85   90   95  100
    *    *    *    *    *    *    *    *    *    *
GCPKGAAGAV TQSLSRTPTA TTSGIRATLT PTVLAPRLPQ PPQNPTNIQN

  105  110  115  120  125  130  135  140  145  150
    *    *    *    *    *    *    *    *    *    *
FQLPPGMVLV RSENGQLLMI PQQALAQMQA QAHAQPQTTM APRPATPTSA

  155  160  165  170  175  180  185  190  195  200
    *    *    *    *    *    *    *    *    *    *
PPVQISTVQA PGTPIIARQV TPTTIIKQVS QAQTTVQPSA TLQRSPGVQP

  205  210  215  220  225  230  235  240  245  250
    *    *    *    *    *    *    *    *    *    *
QLVLGGAAQT ASLGTATAVQ TGTPQRTVPG ATTTSSAATE TMENVKKCKN

  255  260  265  270  275  280  285  290  295  300
    *    *    *    *    *    *    *    *    *    *
FLSTLIKLAS SGKQSTETAA NVKELVQNLL DGKIEAEDFT SRLYRELNSS

  305  310  315  320  325  330  335  340  345  350
    *    *    *    *    *    *    *    *    *    *
PQPYLVPFLK RSLPALRQLT PDSAAFIQQS QQQPPPPTSQ ATTALTAVVL

  355  360  365  370  375  380  385  390  395  400
    *    *    *    *    *    *    *    *    *    *
SSSVQRTAGK TAATVTSALQ PPVLSLTQPT QVGVGKQGQP TPLVIQQPPK

  405  410  415  420  425  430  435  440  445  450
    *    *    *    *    *    *    *    *    *    *
PGALIRPPQV TLTQTPMVAL RQPHNRIMLT TPQQIQLNPL QPVPVVKPAV

  455  460  465  470  475  480  485  490  495  500
    *    *    *    *    *    *    *    *    *    *
LPGTKALSAV SAQAAAAQKN KLKEPGGGSF RDDDDINDVA SMAGVNLSEE

  505  510  515  520  525  530  535  540  545  550
    *    *    *    *    *    *    *    *    *    *
SARILATNSE LVGTLTRSCK DETFLLQAPL QRRILEIGKK HGITELHPDV

  555  560  565  570  575  580  585  590  595  600
    *    *    *    *    *    *    *    *    *    *
VSYVSHATQQ RLQNLVEKIS ETAQQKNFSY KDDDRYEQAS DVRAQLKFFE

  605  610  615  620  625  630  635  640  645  650
    *    *    *    *    *    *    *    *    *    *
QLDQIEKQRK DEQEREILMR AAKSRSRQED PEQLRLKQKA KEMQQQELAQ

  655  660  665  670  675  680  685  690  695  700
    *    *    *    *    *    *    *    *    *    *
MRQRDANLTA LAAIGPRKKR KVDCPGPGSG AEGSGPGSVV PGSSGVGTPR

  705  710  715  720  725  730  735
    *    *    *    *    *    *
QFTRQRITRV NLRDLIFCLE NERETSHSLL LYKAFLK*
```

Sequence Range: 1 to 2152

```
     5     10    15     20     25    30     35     40    45
     *           *           *           *
CTA CTG GCC GTG CTG CAG TTC CTA CGG CAG AGC AAA CTC CGC GAG GCC
Leu Leu Ala Val Leu Gln Phe Leu Arg Gln Ser Lys Leu Arg Glu Ala>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

50     55    60     65     70    75     80     85    90     95
 *           *           *           *           *
GAA GAG GCG CTG CGC CGT GAG GCC GGG CTG CTG GAG GAG GCA GTG GCG
Glu Glu Ala Leu Arg Arg Glu Ala Gly Leu Leu Glu Glu Ala Val Ala>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

  100   105    110    115   120    125    130   135    140
    *           *           *           *           *
GGC TCC GGA GCC CCG GGA GAG GTG GAC AGC GCC GGC GCT GAG GTG ACC
Gly Ser Gly Ala Pro Gly Glu Val Asp Ser Ala Gly Ala Glu Val Thr>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

145   150    155    160   165    170    175   180    185    190
      *           *           *           *           *
AGC GCG CTT CTC AGC CGG GTG ACC GCC TCG GCC CCT GGC CCT GCG GCC
Ser Ala Leu Leu Ser Arg Val Thr Ala Ser Ala Pro Gly Pro Ala Ala>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

195    200    205   210    215    220   225    230    235   240
       *           *           *           *           *
CCC GAC CCT CCG GGC ACT GGC GCT TCG GGG GCC ACG GTC GTC TCA GGT
Pro Asp Pro Pro Gly Thr Gly Ala Ser Gly Ala Thr Val Val Ser Gly>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

   245    250   255    260    265   270    275    280   285
          *           *           *           *
TCA GCC TCA GGT CCT GCG GCT CCG GGT AAA GTT GGA AGT GTT GCT GTG
Ser Ala Ser Gly Pro Ala Ala Pro Gly Lys Val Gly Ser Val Ala Val>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

290    295   300    305    310   315    320    325   330    335
 *           *           *           *           *
GAA GAC CAG CCA GAT GTC AGT GCC GTG TTG TCA GCC TAC AAC CAA CAA
Glu Asp Gln Pro Asp Val Ser Ala Val Leu Ser Ala Tyr Asn Gln Gln>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

  340   345    350    355   360    365    370   375    380
    *           *           *           *           *
GGA GAT CCC ACA ATG TAT GAA GAA TAC TAT AGT GGA CTG AAA CAC TTC
Gly Asp Pro Thr Met Tyr Glu Glu Tyr Tyr Ser Gly Leu Lys His Phe>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

385   390    395    400   405    410    415   420    425    430
      *           *           *           *           *
ATT GAA TGT TCC CTG GAC TGC CAT CGG GCA GAG TTG TCC CAA CTT TTT
Ile Glu Cys Ser Leu Asp Cys His Arg Ala Glu Leu Ser Gln Leu Phe>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

435    440    445   450    455    460   465    470    475   480
       *           *           *           *           *
TAT CCT CTG TTT GTG CAC ATG TAC TTG GAG CTA GTC TAC AAT CAA CAT
Tyr Pro Leu Phe Val His Met Tyr Leu Glu Leu Val Tyr Asn Gln His>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>
```

```
      485     490     495     500     505     510     515     520     525
                *       *       *       *
GAG AAT GAA GCA AAG TCA TTC TTT GAG AAG TTC CAT GGA GAT CAG GAA
Glu Asn Glu Ala Lys Ser Phe Phe Glu Lys Phe His Gly Asp Gln Glu>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  530     535     540     545     550     555     560     565     570     575
    *       *       *       *       *
TGT TAT TAC CAG GAT GAC CTA CGA GTA TTA TCT AGT CTT ACC AAA AAG
Cys Tyr Tyr Gln Asp Asp Leu Arg Val Leu Ser Ser Leu Thr Lys Lys>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

    580     585     590     595     600     605     610     615     620
      *       *       *       *       *       *       *
GAA CAC ATG AAA GGG AAT GAG ACC ATG TTG GAT TTT CGA ACA AGT AAA
Glu His Met Lys Gly Asn Glu Thr Met Leu Asp Phe Arg Thr Ser Lys>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  625     630     635     640     645     650     655     660     665     670
        *       *       *       *       *       *       *
TTT GTT CTG CGT ATT TCC CGT GAC TCG TAC CAA CTC TTG AAG AGG CAT
Phe Val Leu Arg Ile Ser Arg Asp Ser Tyr Gln Leu Leu Lys Arg His>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  675     680     685     690     695     700     705     710     715     720
        *       *       *       *       *       *       *
CTT CAG GAG AAA CAG AAC AAT CAG ATA TGG AAC ATA GTT CAG GAG CAC
Leu Gln Glu Lys Gln Asn Asn Gln Ile Trp Asn Ile Val Gln Glu His>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

    725     730     735     740     745     750     755     760     765
          *       *       *       *
CTC TAC ATT GAC ATC TTT GAT GGG ATG CCG CGT AGT AAG CAA CAG ATA
Leu Tyr Ile Asp Ile Phe Asp Gly Met Pro Arg Ser Lys Gln Gln Ile>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  770     775     780     785     790     795     800     805     810     815
    *       *       *       *       *       *       *
GAT GCG ATG GTG GGA AGT TTG GCA GGA GAG GCT AAA CGA GAG GCA AAC
Asp Ala Met Val Gly Ser Leu Ala Gly Glu Ala Lys Arg Glu Ala Asn>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

    820     825     830     835     840     845     850     855     860
      *       *       *       *       *       *       *
AAA TCA AAG GTA TTT TTT GGT TTA TTA AAA GAA CCA GAA ATT GAG GTA
Lys Ser Lys Val Phe Phe Gly Leu Leu Lys Glu Pro Glu Ile Glu Val>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  865     870     875     880     885     890     895     900     905     910
        *       *       *       *       *       *       *
CCT TTG GAT GAC GAG GAT GAA GAG GGA GAA AAT GAA GAA GGA AAA CCT
Pro Leu Asp Asp Glu Asp Glu Glu Gly Glu Asn Glu Glu Gly Lys Pro>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  915     920     925     930     935     940     945     950     955     960
        *       *       *       *       *       *       *
AAA AAG AAG AAG CCT AAA AAA GAT AGT ATT GGA TCC AAA AGC AAA AAA
Lys Lys Lys Lys Pro Lys Lys Asp Ser Ile Gly Ser Lys Ser Lys Lys>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

    965     970     975     980     985     990     995    1000    1005
          *       *       *       *
```

```
CAA GAT CCC AAT GCT CCA CCT CAG AAC AGA ATC CCT CTT CCT GAG TTG
Gln Asp Pro Asn Ala Pro Pro Gln Asn Arg Ile Pro Leu Pro Glu Leu>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1010   1015  1020   1025   1030  1035   1040   1045  1050   1055
  *             *             *             *             *

AAA GAT TCA GAT AAG TTG GAT AAG ATA ATG AAT ATG AAA GAA ACC ACC
Lys Asp Ser Asp Lys Leu Asp Lys Ile Met Asn Met Lys Glu Thr Thr>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

 1060  1065   1070   1075  1080   1085   1090  1095   1100
    *             *             *             *             *

AAA CGA GTA CGC CTT GGG CCG GAC TGC TTA CCC TCC ATT TGT TTC TAT
Lys Arg Val Arg Leu Gly Pro Asp Cys Leu Pro Ser Ile Cys Phe Tyr>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1105  1110   1115   1120  1125   1130   1135  1140   1145   1150
      *             *             *             *             *

ACA TTT CTC AAT GCT TAC CAG GGT CTC ACT GCA GTG GAT GTC ACT GAT
Thr Phe Leu Asn Ala Tyr Gln Gly Leu Thr Ala Val Asp Val Thr Asp>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1155   1160   1165  1170   1175   1180  1185   1190   1195  1200
       *             *             *             *             *

GAT TCT AGT CTG ATT GCT GGA GGT TTT GCA GAT TCA ACT GTC AGA GTG
Asp Ser Ser Leu Ile Ala Gly Gly Phe Ala Asp Ser Thr Val Arg Val>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

  1205   1210  1215   1220   1225  1230   1235   1240  1245
         *             *             *             *

TGG TCG GTA ACA CCC AAA AAG CTT CGT AGT GTC AAA CAA GCA TCA GAT
Trp Ser Val Thr Pro Lys Lys Leu Arg Ser Val Lys Gln Ala Ser Asp>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1250   1255  1260   1265   1270  1275   1280   1285  1290   1295
  *             *             *             *             *

CTT AGT CTT ATA GAC AAA GAA TCA GAT GAT GTC TTA GAA AGA ATC ATG
Leu Ser Leu Ile Asp Lys Glu Ser Asp Asp Val Leu Glu Arg Ile Met>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

 1300  1305   1310   1315  1320   1325   1330  1335   1340
    *             *             *             *             *

GAT GAG AAA ACA GCA AGT GAG TTG AAG ATT TTG TAT GGT CAC AGT GGG
Asp Glu Lys Thr Ala Ser Glu Leu Lys Ile Leu Tyr Gly His Ser Gly>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1345  1350   1355   1360  1365   1370   1375  1380   1385   1390
      *             *             *             *             *

CCT GTC TAC GGA GCC AGC TTC AGT CCG GAT AGG AAC TAT CTG CTT TCC
Pro Val Tyr Gly Ala Ser Phe Ser Pro Asp Arg Asn Tyr Leu Leu Ser>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

1395   1400   1405  1410   1415   1420  1425   1430   1435  1440
       *             *             *             *             *

TCT TCA GAG GAC GGA ACT GTT AGA TTG TGG AGC CTT CAA ACA TTT ACT
Ser Ser Glu Asp Gly Thr Val Arg Leu Trp Ser Leu Gln Thr Phe Thr>
___a___a___a____TRANSLATION OF HTAF100 DNA (A)_a___a___a___a___>

  1445   1450  1455   1460   1465  1470   1475   1480  1485
         *             *             *             *

TGT TTG GTG GGA TAT AAA GGA CAC AAC TAT CCA GTA TGG GAC ACA CAA
Cys Leu Val Gly Tyr Lys Gly His Asn Tyr Pro Val Trp Asp Thr Gln>
```

```
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

1490   1495  1500   1505   1510  1515   1520   1525  1530   1535
       *           *           *           *           *
TTT TCN CCA TAT GGA TAT TAT TTT GTG TCA GGG GGC CAT GAC CGA GTA
Phe Ser Pro Tyr Gly Tyr Tyr Phe Val Ser Gly Gly His Asp Arg Val>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

 1540  1545   1550   1555  1560   1565   1570  1575   1580
    *           *           *           *           *
GCT CGG CTC TGG GCT ACA GAC CAC TAT CAG CCT TTA AGA ATA TTT GCC
Ala Arg Leu Trp Ala Thr Asp His Tyr Gln Pro Leu Arg Ile Phe Ala>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

1585  1590   1595   1600  1605   1610   1615  1620   1625   1630
       *           *           *           *           *
GGC CAT CTT GCT GAT GTG AAT TGT ACC AGA TTC CAT CCA AAT TCT AAT
Gly His Leu Ala Asp Val Asn Cys Thr Arg Phe His Pro Asn Ser Asn>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

 1635   1640   1645  1650   1655   1660  1665   1670   1675  1680
        *           *           *           *           *
TAT GTT GCT ACG GGC TCT GCA GAC AGA ACT GTG CGG CTC TGG GAC GTC
Tyr Val Ala Thr Gly Ser Ala Asp Arg Thr Val Arg Leu Trp Asp Val>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  1685   1690  1695   1700   1705  1710   1715   1720  1725
         *           *           *           *
CTG AAT GGT AAC TGT GTA AGG ATC TTC ACT GGA CAC AAG GGA CCA ATT
Leu Asn Gly Asn Cys Val Arg Ile Phe Thr Gly His Lys Gly Pro Ile>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

1730   1735  1740   1745   1750  1755   1760   1765  1770   1775
       *           *           *           *           *
CAT TCC TTG ACA TTT TCT CCC AAT GGG AGA TTC CTG GCT ACA GGA GCA
His Ser Leu Thr Phe Ser Pro Asn Gly Arg Phe Leu Ala Thr Gly Ala>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

 1780  1785   1790   1795  1800   1805   1810  1815   1820
    *           *           *           *           *
ACA GAT GGC AGA GTG CTT CTT TGG GAT ATT GGA CAT GGT TTG ATG GTT
Thr Asp Gly Arg Val Leu Leu Trp Asp Ile Gly His Gly Leu Met Val>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

1825  1830   1835   1840  1845   1850   1855  1860   1865   1870
       *           *           *           *           *
GGA GAA TTA AAA GGC CAC ACT GAT ACA GTC TGT TCA CTT AGG TTT AGT
Gly Glu Leu Lys Gly His Thr Asp Thr Val Cys Ser Leu Arg Phe Ser>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

 1875   1880  1885   1890   1895  1900   1905   1910  1915   1920
        *           *           *           *           *
AGA GAT GGT GAA ATT TTG GCA TCA GGT TCA ATG GAT AAT ACA GTT CGA
Arg Asp Gly Glu Ile Leu Ala Ser Gly Ser Met Asp Asn Thr Val Arg>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

  1925   1930  1935   1940   1945  1950   1955   1960  1965
         *           *           *           *
TTA TGG GAT GCT ATC AAA GCC TTT GAA GAT TTA GAG ACC GAT GAC TTT
Leu Trp Asp Ala Ile Lys Ala Phe Glu Asp Leu Glu Thr Asp Asp Phe>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>
```

```
1970   1975  1980   1985   1990  1995   2000   2005  2010   2015
  *            *             *            *             *
ACT ACA GCC ACT GGG CAT ATA AAT TTA CCT GAG AAT TCA CAG GAG TTA
Thr Thr Ala Thr Gly His Ile Asn Leu Pro Glu Asn Ser Gln Glu Leu>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

 2020  2025   2030   2035  2040   2045   2050  2055   2060
    *            *             *            *             *
TTG TTG GGA ACA TAT ATG ACC AAA TCA ACA CCA GTT GTA CAC CTT CAT
Leu Leu Gly Thr Tyr Met Thr Lys Ser Thr Pro Val Val His Leu His>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

2065  2070   2075   2080  2085   2090   2095  2100   2105   2110
    *            *             *            *             *
TTT ACT CGA AGA AAC CTG GTT CTA GCT GCA GGA GCT TAT AGT CCA CAA
Phe Thr Arg Arg Asn Leu Val Leu Ala Ala Gly Ala Tyr Ser Pro Gln>
___a___a___a____TRANSLATION OF HTAF100 DNA [A]_a___a___a___a___>

2115  2120  2125 2130  2135 2140  2145 2150
    *           *            *           *
TAA ACCAT CGGTATTAAA GACCAAAAAA AAAAAAAAAA AA
***>
___>
```

```
     5   10    15   20    25   30    35   40    45   50
     *         *         *         *         *         *
LLAVLQFLRQ SKLREAEEAL RREAGLLEEA VAGSGAPGEV DSAGAEVTSA

    55   60    65   70    75   80    85   90    95  100
     *         *         *         *         *         *
LLSRVTASAP GPAAPDPPGT GASGATVVSG SASGPAAPGK VGSVAVEDQP

   105  110   115  120   125  130   135  140   145  150
     *         *         *         *         *         *
DVSAVLSAYN QQGDPTMYEE YYSGLKHFIE CSLDCHRAEL SQLFYPLFVH

   155  160   165  170   175  180   185  190   195  200
     *         *         *         *         *         *
MYLELVYNQH ENEAKSFFEK FHGDQECYYQ DDLRVLSSLT KKEHMKGNET

   205  210   215  220   225  230   235  240   245  250
     *         *         *         *         *         *
MLDFRTSKFV LRISRDSYQL LKRHLQEKQN NQIWNIVQEH LYIDIFDGMP

   255  260   265  270   275  280   285  290   295  300
     *         *         *         *         *         *
RSKQQIDAMV GSLAGEAKRE ANKSKVFFGL LKEPEIEVPL DDEDEEGENE

   305  310   315  320   325  330   335  340   345  350
     *         *         *         *         *         *
EGKPKKKKPK KDSIGSKSKK QDPNAPPQNR IPLPELKDSD KLDKIMNMKE

   355  360   365  370   375  380   385  390   395  400
     *         *         *         *         *         *
TTKRVRLGPD CLPSICFYTF LNAYQGLTAV DVTDDSSLIA GGFADSTVRV

   405  410   415  420   425  430   435  440   445  450
     *         *         *         *         *         *
WSVTPKKLRS VKQASDLSLI DKESDDVLER IMDEKTASEL KILYGHSGPV

   455  460   465  470   475  480   485  490   495  500
     *         *         *         *         *         *
YGASFSPDRN YLLSSSEDGT VRLWSLQTFT CLVGYKGHNY PVWDTQFSPY

   505  510   515  520   525  530   535  540   545  550
     *         *         *         *         *         *
GYYFVSGGHD RVARLWATDH YQPLRIFAGH LADVNCTRFH PNSNYVATGS

   555  560   565  570   575  580   585  590   595  600
     *         *         *         *         *         *
ADRTVRLWDV LNGNCVRIFT GHKGPIHSLT FSPNGRFLAT GATDGRVLLW

   605  610   615  620   625  630   635  640   645  650
     *         *         *         *         *         *
DIGHGLMVGE LKGHTDTVCS LRFSRDGEIL ASGSMDNTVR LWDAIKAFED

   655  660   665  670   675  680   685  690   695  700
     *         *         *         *         *         *
LETDDFTTAT GHINLPENSQ ELLLGTYMTK STPVVHLHFT RRNLVLAAGA

   705
YSPQ*
```

Sequence Range: 1 to 3820 (SEQ ID NO 19)

```
                10          20          30          40          50
                 *           *           *           *           *
AAT TCC TTT TTT ATA ACA AAC GCA AAT TAG TTA ATT AAA TTC TGG CGC AGA ACC GGC
TTA AGG AAA AAA TAT TGT TTG CGT TTA ATC AAT TAA TTT AAG ACC GCG TCT TGG CCG

                70          80          90          100         110
                 *           *           *           *           *
TGA GCG ATG GAA ACG CAA CCT GAG GTG CCC GAG GTG CCG CTG CGA CCG TTT AAA TTG
ACT CGC TAC CTT TGC GTT GGA CTC CAC GGG CTC CAC GGC GAC GCT GGC AAA TTT AAC
          M   E   T   Q   P   E   V   P   E   V   P   L   R   P   F   K   L

                130         140         150         160         170
                 *           *           *           *           *
CAT CAG GTT GTG AGC CTC ACG GGC ATC AGT TTC GAG CGG AGG AGC ATA ATC GGC GTG
GTA GTC CAA CAC TCG GAG TGC CCG TAG TCA AAG CTC GCC TCC TCG TAT TAG CCG CAC
 H   Q   V   V   S   L   T   G   I   S   F   E   R   R   S   I   I   G   V

                190         200         210         220         230
                 *           *           *           *           *
GAG CTG ACC ATT GTG CCG AAC AGC GAG AAT CTG CGC CTG ATA CGC CTG AAT GCC AAG
CTC GAC TGG TAA CAC GGC TTG TCG CTC TTA GAC GCG GAC TAT GCG GAC TTA CGG TTC
 E   L   T   I   V   P   N   S   E   N   L   R   L   I   R   L   N   A   K

                250         260         270         280         290
                 *           *           *           *           *
CTG AGA ATC TAC AGC GTC GTT TTG AAC GAT GTC TGC CAG GCG GAT TTC ACG TAC TTC
GAC TCT TAG ATG TCG CAG CAA AAC TTG CTA CAG ACG GTC CGC CTA AAG TGC ATG AAG
 L   R   I   Y   S   V   V   L   N   D   V   C   Q   A   D   F   T   Y   F

                310         320         330         340         350
                 *           *           *           *           *
CCC TTC CAG AAC ATC TGC TAC AAG GAG CCC AAG AGC CGC GCT CTG GAG GTC TAC TCC
GGG AAG GTC TTG TAG ACG ATG TTC CTC GGG TTC TCG GCG CGA GAC CTC CAG ATG AGG
 P   F   Q   N   I   C   Y   K   E   P   K   S   R   A   L   E   V   Y   S

                370         380         390         400         410
                 *           *           *           *           *
CAT CAT CTG ACC GCC GCC CAG TAC ACC GAT CCC GAT GTG AAC AAC GGC GAA CTG CTC
GTA GTA GAC TGG CGG CGG GTC ATG TGG CTA GGG CTA CAC TTG TTG CCG CTT GAC GAG
 H   H   L   T   A   A   Q   Y   T   D   P   D   V   N   N   G   E   L   L

                430         440         450         460         470
                 *           *           *           *           *
CAG GTT CCG CCC GAG GGC TAC TCT ATG ATC CAG GAG GGT CAG GGT CTG CGC ATC CGC
GTC CAA GGC GGG CTC CCG ATG AGA TAC TAG GTC CTC CCA GTC CCA GAC GCG TAG GCG
 Q   V   P   P   E   G   Y   S   M   I   Q   E   G   Q   G   L   R   I   R

                490         500         510         520         530
                 *           *           *           *           *
GAG TTC TCG TTG GAG AAT CCC AAA TGC GGC GTA CAT TTT GTC ATA CCA CCC GCT TCA
CTC AAG AGC AAC CTC TTA GGG TTT ACG CCG CAT GTA AAA CAG TAT GGT GGG CGA AGT
 E   F   S   L   E   N   P   K   C   G   V   H   F   V   I   P   P   A   S

                550         560         570         580         590
                 *           *           *           *           *
GAC GAG GAG ACA CAG ATG AAC AGC TCG CAT ATG TTC ACC AAT TGC TAT GAA AAC TCG
CTG CTC CTC TGT GTC TAC TTG TCG AGC GTA TAC AAG TGG TTA ACG ATA CTT TTG AGC
 D   E   E   T   Q   M   N   S   S   H   M   F   T   N   C   Y   E   N   S

                610         620         630         640         650
```

FIG 10 (8 pg)

```
                *           *           *           *           *
AGA TTG TGG TTT CCC TGC GTG GAC AGT TTC GCC GAT CCC TGC ACC TGG CGG CTG GAG
TCT AAC ACC AAA GGG ACG CAC CTG TCA AAG CGG CTA GGG ACG TGG ACC GCC GAC CTC
 R   L   W   F   P   C   V   D   S   F   A   D   P   C   T   W   R   L   E

          670         680         690         700         710
            *           *           *           *           *
ACT GTC GAC AAA AAT ATG ACC GCC GTT TCG TGT GGA GAA CTT CTA GAA GTC ATT ATG
TGA CAG CTG TTT TTA TAC TGG CGG CAA AGC ACA CCT CTT GAA GAT CTT CAG TAA TAC
 T   V   D   K   N   M   T   A   V   S   C   G   E   L   L   E   V   I   M

          730         740         750         760         770
            *           *           *           *           *
CCA GAT CTG CGA AAG AAA ACC TTC CAC TAT TCG GTT AGC ACA CCA GTA TGT GCA CCA
GGT CTA GAC GCT TTC TTT TGG AAG GTG ATA AGC CAA TCG TGT GGT CAT ACA CGT GGT
 P   D   L   R   K   K   T   F   H   Y   S   V   S   T   P   V   C   A   P

          790         800         810         820         830
            *           *           *           *           *
ATT GCG CTG GCT GTG GGT CAG TTT GAG ATC TAC GTG GAT CCG CAC ATG CAT GAA GTG
TAA CGC GAC CGA CAC CCA GTC AAA CTC TAG ATG CAC CTA GGC GTG TAC GTA CTT CAC
 I   A   L   A   V   G   Q   F   E   I   Y   V   D   P   H   M   H   E   V

          850         860         870         880         890
            *           *           *           *           *
CAC TTT TGT CTG CCC GGA TTG TTG CCG CTG TTA AAA AAT ACG GTT CGC TAT TTG CAC
GTG AAA ACA GAC GGG CCT AAC AAC GGC GAC AAT TTT TTA TGC CAA GCG ATA AAC GTG
 H   F   C   L   P   G   L   L   P   L   L   K   N   T   V   R   Y   L   H

          910         920         930         940         950
            *           *           *           *           *
GCA TTT GAA TTT TAC GAG GAG ACC TTA TCT ACG CGC TAC CCA TTC AGT TGC TAC AAA
CGT AAA CTT AAA ATG CTC CTC TGG AAT AGA TGC GCG ATG GGT AAG TCA ACG ATG TTT
 A   F   E   F   Y   E   E   T   L   S   T   R   Y   P   F   S   C   Y   K

          970         980         990         1000        1010
            *           *           *           *           *
GTG TTT GTA GAC GAA TTG GAC ACG GAC ATA AGT GCC TAT GCC ACT ATG AGC ATT GCT
CAC AAA CAT CTG CTT AAC CTG TGC CTG TAT TCA CGG ATA CGG TGA TAC TCG TAA CGA
 V   F   V   D   E   L   D   T   D   I   S   A   Y   A   T   M   S   I   A

          1030        1040        1050        1060        1070
            *           *           *           *           *
GTG AAC CTG CTG CAC TCC ATA GCT ATC ATC GAT CAG ACC TAT ATA TCT CGA ACC TTT
CAC TTG GAC GAC GTG AGG TAT CGA TAG TAG CTA GTC TGG ATA TAT AGA GCT TGG AAA
 V   N   L   L   H   S   I   A   I   I   D   Q   T   Y   I   S   R   T   F

          1090        1100        1110        1120        1130
            *           *           *           *           *
TCG CGC GCT GTG GCT GAG CAA TTC TTC GGC TGC TTT ATT ACA TCG CAT CAT TGG TCG
AGC GCG CGA CAC CGA CTC GTT AAG AAG CCG ACG AAA TAA TGT AGC GTA GTA ACC AGC
 S   R   A   V   A   E   Q   F   F   G   C   F   I   T   S   H   H   W   S

          1150        1160        1170        1180        1190
            *           *           *           *           *
ACC TGG CTG GCC AAG GGC ATT GCG GAG TAC CTG TGT GGA TTG TAT TCC AGG AAG TGC
TGG ACC GAC CGG TTC CCG TAA CGC CTC ATG GAC ACA CCT AAC ATA AGG TCC TTC ACG
 T   W   L   A   K   G   I   A   E   Y   L   C   G   L   Y   S   R   K   C

          1210        1220        1230        1240        1250
            *           *           *           *           *
GGC AAC AAC GAG TAC CGT GCT TGG GTG CAA TCT GAA CTG GCG CGT GTC GTT CGC TAC
```

```
CCG TTG TTG CTC ATG GCA CGA ACC CAC GTT AGA CTT GAC CGC GCA CAG CAA GCG ATG
 G   N   N   E   Y   R   A   W   V   Q   S   E   L   A   R   V   V   R   Y

          1270        1280        1290        1300        1310
           *           *           *           *           *
GAG CAG TAT GGC GGC ATT ATT CTC GAT TGC AGT CAG CCG CCA GCA CCT TTG CCT GTT
CTC GTC ATA CCG CCG TAA TAA GAG CTA ACG TCA GTC GGC GGT CGT GGA AAC GGA CAA
 E   Q   Y   G   G   I   I   L   D   C   S   Q   P   P   A   P   L   P   V

          1330        1340        1350        1360        1370
           *           *           *           *           *
GGC ACA AAT CAA TCG GCT GCT TCC AGC AAA CAG CAG GAG ATT GTC CAC TAT TTT CCC
CCG TGT TTA GTT AGC CGA CGA AGG TCG TTT GTC GTC CTC TAA CAG GTG ATA AAA GGG
 G   T   N   Q   S   A   A   S   S   K   Q   Q   E   I   V   H   Y   F   P

          1390        1400        1410        1420        1430
           *           *           *           *           *
AAG AGT TTG CAC ACC GTA TCG CCG AAG TAT GTG GAG GCG ATG CGA AGG AAA GCG CAT
TTC TCA AAC GTG TGG CAT AGC GGC TTC ATA CAC CTC CGC TAC GCT TCC TTT CGC GTA
 K   S   L   H   T   V   S   P   K   Y   V   E   A   M   R   R   K   A   H

          1450        1460        1470        1480        1490
           *           *           *           *           *
GTA ATC CGA ATG CTG GAG AAC CGC ATC GGG CAG GAG CTG CTG ATT CAG GTG TTC AAT
CAT TAG GCT TAC GAC CTC TTG GCG TAG CCC GTC CTC GAC GAC TAA GTC CAC AAG TTA
 V   I   R   M   L   E   N   R   I   G   Q   E   L   L   I   Q   V   F   N

          1510        1520        1530        1540        1550
           *           *           *           *           *
CAA TTG GCT TTG GCT TCT AGT GCG GCA ACG ACG AAG ATC GGT GCA GGA CTC TGG TCT
GTT AAC CGA AAC CGA AGA TCA CGC CGT TGC TGC TTC TAG CCA CGT CCT GAG ACC AGA
 Q   L   A   L   A   S   S   A   A   T   T   K   I   G   A   G   L   W   S

          1570        1580        1590        1600        1610
           *           *           *           *           *
CTG CTC ATC TCG AAC CAA CAT TTT TAT CAA GGC CAT CTT CAC GTA ACC GGA AAA GAT
GAC GAG TAG AGC TTG GTT GTA AAA ATA GTT CCG GTA GAA GTG CAT TGG CCT TTT CTA
 L   L   I   S   N   Q   H   F   Y   Q   G   H   L   H   V   T   G   K   D

          1630        1640        1650        1660        1670
           *           *           *           *           *
TCT GTC TTC ATG GAC CAG TGG GTG CGC ACT GGA GGG CAC GCC AAG TTT TCG CTC ACA
AGA CAG AAG TAC CTG GTC ACC CAC GCG TGA CCT CCC GTG CGG TTC AAA AGC GAG TGT
 S   V   F   M   D   Q   W   V   R   T   G   G   H   A   K   F   S   L   T

          1690        1700        1710        1720        1730
           *           *           *           *           *
GTG TTC AAT CGC AAG AGA AAC ACG ATT GAA CTG GAA ATC CGC CAG GAC TAT GTT AAT
CAC AAG TTA GCG TTC TCT TTG TGC TAA CTT GAC CTT TAG GCG GTC CTG ATA CAA TTA
 V   F   N   R   K   R   N   T   I   E   L   E   I   R   Q   D   Y   V   N

          1750        1760        1770        1780        1790
           *           *           *           *           *
CGG GGA ATT AGA AAA TAC AAT GGT CCA TTG ATG GTG CAG CTG CAG GAG TTG GAT GGA
GCC CCT TAA TCT TTT ATG TTA CCA GGT AAC TAC CAC GTC GAC GTC CTC AAC CTA CCT
 R   G   I   R   K   Y   N   G   P   L   M   V   Q   L   Q   E   L   D   G

          1810        1820        1830        1840        1850
           *           *           *           *           *
TTT AAG CAC ACA TTG CAG ATT GAG AGT ACC CTG GTA AAG TCC GAT ATC ACT TGT CAC
AAA TTC GTG TGT AAC GTC TAA CTC TCA TGG GAC CAT TTC AGG CTA TAG TGA ACA GTG
 F   K   H   T   L   Q   I   E   S   T   L   V   K   S   D   I   T   C   H
```

```
          1870        1880        1890        1900        1910
                *           *           *           *           *
AAG AGC AGG CGT AAC AAA AAG AAG AAG ATC CCC TTG TGC ACC GGT GAG GAA GTG GAT
TTC TCG TCC GCA TTG TTT TTC TTC TTC TAG GGG AAC ACG TGG CCA CTC CTT CAC CTA
 K   S   R   R   N   K   K   K   K   I   P   L   C   T   G   E   E   V   D

          1930        1940        1950        1960        1970
                *           *           *           *           *
GAT TTA TCA GCC ATG GAC GAC TCA CCT GTG CTT TGG ATC CGC CTC GAT CCC GAA ATG
CTA AAT AGT CGG TAC CTG CTG AGT GGA CAC GAA ACC TAG GCG GAG CTA GGG CTT TAC
 D   L   S   A   M   D   D   S   P   V   L   W   I   R   L   D   P   E   M

          1990        2000        2010        2020        2030
                *           *           *           *           *
CTG CTG CGC GAC CTC ATA ATC GAA CAG CCC GAC TTC CAG TGG CAG TAT CAG CTT CGG
GAC GAC GCG CTG GAG TAT TAG CTT GTC GGG CTG AAG GTC ACC GTC ATA GTC GAA GCC
 L   L   R   D   L   I   I   E   Q   P   D   F   Q   W   Q   Y   Q   L   R

          2050        2060        2070        2080        2090
                *           *           *           *           *
GAA CGT GAT GTT ACT GCT CAA TTT CAG GCG ATT CAA GCC CTG CAA AAG TAC CCC ACG
CTT GCA CTA CAA TGA CGA GTT AAA GTC CGC TAA GTT CGG GAC GTT TTC ATG GGG TGC
 E   R   D   V   T   A   Q   F   Q   A   I   Q   A   L   Q   K   Y   P   T

          2110        2120        2130        2140        2150
                *           *           *           *           *
GCC ACC AGG CTT GCT TTA ACC GAC ACC ATA GAA AGC GAA CGT TGC TTC TAT CAG GTG
CGG TGG TCC GAA CGA AAT TGG CTG TGG TAT CTT TCG CTT GCA ACG AAG ATA GTC CAC
 A   T   R   L   A   L   T   D   T   I   E   S   E   R   C   F   Y   Q   V

          2170        2180        2190        2200        2210
                *           *           *           *           *
TGC GAG GCA GCC CAC AGC TTG ACC AAA GTG GCC AAC CAG ATG GTG GCC TCC TGG AGT
ACG CTC CGT CGG GTG TCG AAC TGG TTT CAC CGG TTG GTC TAC CAC CGG AGG ACC TCA
 C   E   A   A   H   S   L   T   K   V   A   N   Q   M   V   A   S   W   S

          2230        2240        2250        2260        2270
                *           *           *           *           *
CCG CCC GCC ATG CTG AAC ATA TTT AGG AAG TTT TTC GGC TCA TTT AGT GCT CCG CAC
GGC GGG CGG TAC GAC TTG TAT AAA TCC TTC AAA AAG CCG AGT AAA TCA CGA GGC GTG
 P   P   A   M   L   N   I   F   R   K   F   F   G   S   F   S   A   P   H

          2290        2300        2310        2320        2330
                *           *           *           *           *
ATC AAA CTG AAC AAC TTC TCC AAC TTT CAG CTG TAC TTC CTG CAG AAG GCT ATT CCC
TAG TTT GAC TTG TTG AAG AGG TTG AAA GTC GAC ATG AAG GAC GTC TTC CGA TAA GGG
 I   K   L   N   N   F   S   N   F   Q   L   Y   F   L   Q   K   A   I   P

          2350        2360        2370        2380        2390
                *           *           *           *           *
GCC ATG GCA GGT CTG CGC ACA TCT CAT GGT ATT TGC CCG CCG GAA GTG ATG CGT TTT
CGG TAC CGT CCA GAC GCG TGT AGA GTA CCA TAA ACG GGC GGC CTT CAC TAC GCA AAA
 A   M   A   G   L   R   T   S   H   G   I   C   P   P   E   V   M   R   F

          2410        2420        2430        2440        2450
                *           *           *           *           *
TTC GAT CTC TTC AAG TAC AAC GAG AAT TCG CGT AAC CAT TAC ACG GAT GCA TAC TAC
AAG CTA GAG AAG TTC ATG TTG CTC TTA AGC GCA TTG GTA ATG TGC CTA CGT ATG ATG
 F   D   L   F   K   Y   N   E   N   S   R   N   H   Y   T   D   A   Y   Y

          2470        2480        2490        2500        2510
```

```
                *           *           *           *           *
GCA GCT TTG GTA GAA GCT CTA GGC GAA ACC TTA ACA CCT GTG GTC TCC GTT GCT ATC
CGT CGA AAC CAT CTT CGA GAT CCG CTT TGG AAT TGT GGA CAC CAG AGG CAA CGA TAG
 A   A   L   V   E   A   L   G   E   T   L   T   P   V   V   S   V   A   I

        2530        2540        2550        2560        2570
           *           *           *           *           *
GGC ACA CAA ATC ACT ACG GAC AGT CTA TCC ACG GAT GCG AAA CTT GTG CTA GAT GAA
CCG TGT GTT TAG TGA TGC CTG TCA GAT AGG TGC CTA CGC TTT GAA CAC GAT CTA CTT
 G   T   Q   I   T   T   D   S   L   S   T   D   A   K   L   V   L   D   E

        2590        2600        2610        2620        2630
           *           *           *           *           *
ACA CGT CTG CTG AAC ATG GAG AAA CAT CTA CCC TCG TAC AAG TAC ATG GTG TCC GTG
TGT GCA GAC GAC TTG TAC CTC TTT GTA GAT GGG AGC ATG TTC ATG TAC CAC AGG CAC
 T   R   L   L   N   M   E   K   H   L   P   S   Y   K   Y   M   V   S   V

        2650        2660        2670        2680        2690
           *           *           *           *           *
TGT CTG AAG GTC ATC CGG AAG CTG CAA AAA TTC GGT CAT CTG CCC TCA CTG CCG CAC
ACA GAC TTC CAG TAG GCC TTC GAC GTT TTT AAG CCA GTA GAC GGG AGT GAC GGC GTG
 C   L   K   V   I   R   K   L   Q   K   F   G   H   L   P   S   L   P   H

        2710        2720        2730        2740        2750
           *           *           *           *           *
TAC CGC AGC TAT GCC GAA TAT GGA ATA TAT CTC GAT CTC CGC ATT GCT GCT ATG GAG
ATG GCG TCG ATA CGG CTT ATA CCT TAT ATA GAG CTA GAG GCG TAA CGA CGA TAC CTC
 Y   R   S   Y   A   E   Y   G   I   Y   L   D   L   R   I   A   A   M   E

        2770        2780        2790        2800        2810
           *           *           *           *           *
CTC GTG GAC TTT GTG AAA GTG GAT GGG CGC AGC GAG GAT TTG GAA CAT TTG ATT ACT
GAG CAC CTG AAA CAC TTT CAC CTA CCC GCG TCG CTC CTA AAC CTT GTA AAC TAA TGA
 L   V   D   F   V   K   V   D   G   R   S   E   D   L   E   H   L   I   T

        2830        2840        2850        2860        2870
           *           *           *           *           *
CTG GAA ACT GAT CCG GAT CCG GCT GCT CGC CAT GCA CTG GCC CAA CTG CTG ATC GAT
GAC CTT TGA CTA GGC CTA GGC CGA CGA GCG GTA CGT GAC CGG GTT GAC GAC TAG CTA
 L   E   T   D   P   D   P   A   A   R   H   A   L   A   Q   L   L   I   D

        2890        2900        2910        2920        2930
           *           *           *           *           *
CCG CCT TTC ACA CGC GAA TCT CGC AGC CGT CTG GAT AAA CCC AAT CTC GTG GAT CGT
GGC GGA AAG TGT GCG CTT AGA GCG TCG GCA GAC CTA TTT GGG TTA GAG CAC CTA GCA
 P   P   F   T   R   E   S   R   S   R   L   D   K   P   N   L   V   D   R

        2950        2960        2970        2980        2990
           *           *           *           *           *
TGG TTC AGT ATT AAT CGC TTG CCC TAC GAT ACC CAA STG CGC TGC GAT ATT GTC GAT
ACC AAG TCA TAA TTA GCG AAC GGG ATG CTA TGG GTT SAC GCG ACG CTA TAA CAG CTA
 W   F   S   I   N   R   L   P   Y   D   T   Q   X   R   C   D   I   V   D

        3010        3020        3030        3040        3050
           *           *           *           *           *
TAC TAC GCA CTG TAC GGA ACT AAG CGT CCG AAT TGC TTG CAG GCC GGC GAG AAC CAA
ATG ATG CGT GAC ATG CCT TGA TTC GCA GGC TTA ACG AAC GTC CGG CCG CTC TTG GTT
 Y   Y   A   L   Y   G   T   K   R   P   N   C   L   Q   A   G   E   N   Q

        3070        3080        3090        3100        3110
           *           *           *           *           *
TTC TAC AAG GAT TTG ATG AAG GAC AAT AAT AGC AGT GTA GGC AGC GTA ACC GGC AGC
```

```
AAG ATG TTC CTA AAC TAC TTC CTG TTA TTA TCG TCA CAT CCG TCG CAT TGG CCG TCG
 F   Y   K   D   L   M   K   D   N   N   S   S   V   G   S   V   T   G   S

          3130        3140        3150        3160        3170
             *           *           *           *           *
AAG AAG ACC AGT GAT TCA AAG TCA CAT TTG CCA ACA CCA ACG AAT ACT TTG GAC AAT
TTC TTC TGG TCA CTA AGT TTC AGT GTA AAC GGT TGT GGT TGC TTA TGA AAC CTG TTA
 K   K   T   S   D   S   K   S   H   L   P   T   P   T   N   T   L   D   N

          3190        3200        3210        3220        3230
             *           *           *           *           *
CCA CAG GAG CGG CAA AAG CCG GCA ATG GTT ACC ATC AAG CGA ACG GCC ACA GAA GCA
GGT GTC CTC GCC GTT TTC GGC CGT TAC CAA TGG TAG TTC GCT TGC CGG TGT CTT CGT
 P   Q   E   R   Q   K   P   A   M   V   T   I   K   R   T   A   T   E   A

          3250        3260        3270        3280        3290
             *           *           *           *           *
GAG GTG GGC GAT GAG ATT ATC AAG CTG GAA CGC AGC GAG GAG ATC ACC GTG CTA GAT
CTC CAC CCG CTA CTC TAA TAG TTC GAC CTT GCG TCG CTC CTC TAG TGG CAC GAT CTA
 E   V   G   D   E   I   I   K   L   E   R   S   E   E   I   T   V   L   D

          3310        3320        3330        3340        3350
             *           *           *           *           *
CCA GTT AAC GTG CAG GCC TAT GAC AGT GAG ACC AAA GTG AAT GCC CTG CAG GCA GAT
GGT CAA TTG CAC GTC CGG ATA CTG TCA CTC TGG TTT CAC TTA CGG GAC GTC CGT CTA
 P   V   N   V   Q   A   Y   D   S   E   T   K   V   N   A   L   Q   A   D

          3370        3380        3390        3400        3410
             *           *           *           *           *
GAA GCA CGT GAT ACC CAT CAG GCT GCC AAG CGC CTT AAG AAC GAA ATG TAC GCC GAG
CTT CGT GCA CTA TGG GTA GTC CGA CGG TTC GCG GAA TTC TTG CTT TAC ATG CGG CTC
 E   A   R   D   T   H   Q   A   A   K   R   L   K   N   E   M   Y   A   E

          3430        3440        3450        3460        3470
             *           *           *           *           *
GAT AAC TCA TCC ACA ATG CTC GAC GTG GGC GAC TCC ACC AGA TAT GAG AGT AGC CAC
CTA TTG AGT AGG TGT TAC GAG CTG CAC CCG CTG AGG TGG TCT ATA CTC TCA TCG GTG
 D   N   S   S   T   M   L   D   V   G   D   S   T   R   Y   E   S   S   H

          3490        3500        3510        3520        3530
             *           *           *           *           *
GAG GGC AAA TTG AAG TCC GGC GAT GGT GGG CTC AAG AAG AAA AAG AAG AAG GAG AAG
CTC CCG TTT AAC TTC AGG CCG CTA CCA CCC GAG TTC TTC TTT TTC TTC TTC CTC TTC
 E   G   K   L   K   S   G   D   G   G   L   K   K   K   K   K   K   E   K

          3550        3560        3570        3580        3590
             *           *           *           *           *
AAG CAT AAG CAC AAA CVC AAG CAT AGG CAC AGC AAG GAC AAG GAC AAG GAG CGA AAG
TTC GTA TTC GTG TTT GBG TTC GTA TCC GTG TCG TTC CTG TTC CTG TTC CTC GCT TTC
 K   H   K   H   K   X   K   H   R   H   S   K   D   K   D   K   E   R   K

          3610        3620        3630        3640        3650
             *           *           *           *           *
AAG GAC AAG CGT GAC CCG CAT ATA TTC ACC CTG CAG GCG CGC GAG ACA GCC ACT CCG
TTC CTG TTC GCA CTG GGC GTA TAT AAG TGG GAC GTC CGC GCG CTC TGT CGG TGA GGC
 K   D   K   R   D   P   H   I   F   T   L   Q   A   R   E   T   A   T   P

          3670        3680        3690        3700        3710
             *           *           *           *           *
ACT CTC AGC TCG GAG GAC AGT AGC AAC AGC AAT AGC CTG CCG CCC ATG AAC CTT AAC
TGA GAG TCG AGC CTC CTG TCA TCG TTG TCG TTA TCG GAC GGC GGG TAC TTG GAA TTG
 T   L   S   S   E   D   S   S   N   S   N   S   L   P   P   M   N   L   N
```

```
          3730          3740          3750          3760          3770
            *             *             *             *             *
GTG AGG GTT CCT ACA GGT GGG GAA ATT GCA ATG TTT GGG GGA TAG ATG ACA GAA TAA
CAC TCC CAA GGA TGT CCA CCC CTT TAA CGT TAC AAA CCC CCT ATC TAC TGT CTT ATT
 V   R   V   P   T   G   G   E   I   A   M   F   G   G   *   M   T   E   *

          3790          3800          3810          3820
            *             *             *             *
TAT AAT ACC TTA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA A
ATA TTA TGG AAT TTT TTT TTT TTT TTT TTT TTT TTT TTT T
 Y   N   T   L   K   K   K   K   K   K   K   K   K   X>
```

LOCUS TRANSLDTAF 1217 AA PROT

FEATURES	From	To/Span	Description
Peptide	1	> 1217	67 to 3820 of dTAF150 (translated) [Split]
	< 1218	1218	67 to 3820 of dTAF150 (translated) [Split]

```
   1 METQPEVPEV PLRPFKLAHQ VVSLTGISFE RRSIIGVVEL TIVPNSENLR LIRLNAKQLR
  61 IYSVVLNDVC QADFTYFDPF QNICYKEPKS RALEVYSKHH LTAAQYTDPD VNNGELLIQV
 121 PPEGYSMIQE GQGLRIRIEF SLENPKCGVH FVIPPASTDE ETQMNSSHMF TNCYENSSRL
 181 WFPCVDSFAD PCTWRLEFTV DKNMTAVSCG ELLEVIMTPD LRKKTFHYSV STPVCAPNIA
 241 LAVGQFEIYV DPHMHEVTHF CLPGLLPLLK NTVRYLHEAF EFYEETLSTR YPFSCYKQVF
 301 VDELDTDISA YATMSIASVN LLHSIAIIDQ TYISRTFMSR AVAEQFFGCF ITSHHWSDTW
 361 LAKGIAEYLC GLYSRKCFGN NEYRAWVQSE LARVVRYEEQ YGGIILDCSQ PPAPLPVSGT
 421 NQSAASSKQQ EIVHYFPIKS LHTVSPKYVE AMRRKAHFVI RMLENRIGQE LLIQVFNKQL
 481 ALASSAATTK IGAGLWSQLL ISNQHFYQGH LHVTGKDMSV FMDQWVRTGG HAKFSLTSVF
 541 NRKRNTIELE IRQDYVNQRG IRKYNGPLMV QLQELDGTFK HTLQIESTLV KSDITCHSKS
 601 RRNKKKKIPL CTGEEVDMDL SAMDDSPVLW IRLDPEMILL RDLIIEQPDF QWQYQLRHER
 661 DVTAQFQAIQ ALQKYPTNAT RLALTDTIES ERCFYQVRCE AAHSLTKVAN QMVASWSGPP
 721 AMLNIFRKFF GSFSAPHIIK LNNFSNFQLY FLQKAIPVAM AGLRTSHGIC PPEVMRFLFD
 781 LFKYNENSRN HYTDAYYRAA LVEALGETLT PVVSVAIHGT QITTDSLSTD AKLVLDEVTR
 841 LLNMEKHLPS YKYMVSVSCL KVIRKLQKFG HLPSLPHIYR SYAEYGIYLD LRIAAMECLV
 901 DFVKVDGRSE DLEHLITLLE TDPDPAARHA LAQLLIDNPP FTRESRSRLD KPNLVDRLWF
 961 SINRLPYDTQ XRCDIVDLYY ALYGTKRPNC LQAGENQSFY KDLMKDNNSS VGSVTGSFKK
1021 TSDSKSHLPT PTNTLDNEPQ ERQKPAMVTI KRTATEAFEV GDEIIKLERS EEITVLDEPV
1081 NVQAYDSETK VNALQADEEA RDTHQAAKRL KNEMYAEDDN SSTMLDVGDS TRYESSHEEG
1141 KLKSGDGGLK KKKKKEKKKH KHKXKHRHSK DKDKERKDKD KRDPHIFTLQ ARETATPDTL
1201 SSEDSSNSNS LPPMNLN
```

Sequence Range: 1 to 872

```
          10         20         30         40         50         60
           *          *          *          *          *          *
CCAAAAATCC GCCCAACTTA CTGTACTTTC CCCAAACACT TCCAACCAAC CGACCTACCA
GGTTTTTAGG CGGGTTGAAT GACATGAAAG GGGTTTGTGA AGGTTGGTTG GCTGGATGGT

          70         80         90         100        110
           *          *          *          *          *
CCCACTTGAT TTGACTCTGA AGAAACCCAA AAGCA ATG TCG GAT CTC TTT ACC ACT
GGGTGAACTA AACTGAGACT TCTTTGGGTT TTCGT TAC AGC CTA GAG AAA TGG TGA
                                       M   S   D   L   F   T   T>
                                     TRANSLATION OF 802 F     >

   120        130        140        150        160
    *          *          *          *          *
TTC GAT AGC AAC GGC GTC GCG AGG CAC CAC CTG CAC CAC AAC CAC AAC
AAG CTA TCG TTG CCG CAG CGC TCC GTG GTG GAC GTG GTG TTG GTG TTG
 F   D   S   N   G   V   A   R   H   H   L   H   H   N   H   N>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

    170        180        190        200        210
     *          *          *          *          *
TCC ACA TCG TCC GCC AGC GGA CTG CTC CAC GAC CCA CCC ATG GCC TCG
AGG TGT AGC AGG CGG TCG CCT GAC GAG GTG CTG GGT GGG TAC CGG AGC
 S   T   S   S   A   S   G   L   L   H   D   P   P   M   A   S>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

       220        230        240        250        260
        *          *          *          *          *
CCC TCC CAG CAC AGT CCG ATG ACC AAC AAC AGC AAC TCA TCC TCG CAG
GGG AGG GTC GTG TCA GGC TAC TGG TTG TTG TCG TTG AGT AGG AGC GTC
 P   S   Q   H   S   P   M   T   N   N   S   N   S   S   S   Q>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

          270        280        290        300
           *          *          *          *
AAC GGC GGA CCG GTT TCC GGT TTG GGT ACG GGA ACG GGC CCC ATA TCT
TTG CCG CCT GGC CAA AGG CCA AAC CCA TGC CCT TGC CCG GGG TAT AGA
 N   G   G   P   V   S   G   L   G   T   G   T   G   P   I   S>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

310        320        330        340        350
 *          *          *          *          *
GGT GGT AGC AAG TCA TCC AAT CAC ACA TCA TCC GCC GCC GGT TCC GAG
CCA CCA TCG TTC AGT AGG TTA GTG TGT AGT AGG CGG CGG CCA AGG CTC
 G   G   S   K   S   S   N   H   T   S   S   A   A   G   S   E>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

   360        370        380        390        400
    *          *          *          *          *
AAC ACT CCC ATG CTT ACC AAA CCG CGT CTC ACA GAG CTC GTC CGA GAG
TTG TGA GGG TAC GAA TGG TTT GGC GCA GAG TGT CTC GAG CAG GCT CTC
 N   T   P   M   L   T   K   P   R   L   T   E   L   V   R   E>
   a   a   a   a TRANSLATION OF 802 FULL (A)   a   a   a   a   >

    410        420        430        440        450
     *          *          *          *          *
GTG GAT ACC ACC ACG CAG CTG GAC GAG GAT GTT GAG GAG CTT CTG CTT
CAC CTA TGG TGG TGC GTC GAC CTG CTC CTA CAA CTC CTC GAA GAC GAA
 V   D   T   T   T   Q   L   D   E   D   V   E   E   L   L   L>
___a___a___a___a_TRANSLATION OF 802 FULL (A)___a___a___a___a___>
```


FIG 8 (2 Pg)

```
        460          470          480          490          500
         *            *            *            *            *
CAG ATC ATC GAC GAC TTT GTG AGG GAC ACC GTC AAG TCG ACG AGC GCC
GTC TAG TAG CTG CTG AAA CAC TCC CTG TGG CAG TTC AGC TGC TCG CGG
 Q   I   I   D   D   F   V   R   D   T   V   K   S   T   S   A>
   a   a   a   a TRANSLATION OF 802 FULL [A]   a   a   a   a   >

          510          520          530          540
           *            *            *            *
TTC GCC AAG CAC CGA AAG TCT AAC AAG ATC GAG GTG CGC GAC GTG CAG
AAG CGG TTC GTG GCT TTC AGA TTG TTC TAG CTC CAC GCG CTG CAC GTC
 F   A   K   H   R   K   S   N   K   I   E   V   R   D   V   Q>
   a   a   a   a TRANSLATION OF 802 FULL [A]   a   a   a   a   >

550          560          570          580          590
 *            *            *            *            *
CTG CAC TTT GAG CGG AAG TAC AAC ATG TGG ATA CCC GGC TTC GGT ACG
GAC GTG AAA CTC GCC TTC ATG TTG TAC ACC TAT GGG CCG AAG CCA TGC
 L   H   F   E   R   K   Y   N   M   W   I   P   G   F   G   T>
   a   a   a   a TRANSLATION OF 802 FULL [A]   a   a   a   a   >

  600          610          620          630          640
   *            *            *            *            *
GAC GAA CTG CGT CCC TAC AAG CGG GCA GCT GTC ACG GAG GCG CAC AAA
CTG CTT GAC GCA GGG ATG TTC GCC CGT CGA CAG TGC CTC CGC GTG TTT
 D   E   L   R   P   Y   K   R   A   A   V   T   E   A   H   K>
   a   a   a   a TRANSLATION OF 802 FULL [A]   a   a   a   a   >

    650          660          670          680          690
     *            *            *            *            *
CAG CGC CTT GCC CTC ATA CGG AAA ACG ATC AAG AAA TAC TAG AGGA
GTC GCG GAA CGG GAG TAT GCC TTT TGC TAG TTC TTT ATG ATC TCCT
 Q   R   L   A   L   I   R   K   T   I   K   K   Y   *>
   a   a   a TRANSLATION OF 802 FULL [A]   a   a   a   >

        700          710          720          730          740          750
         *            *            *            *            *            *
TTGGATCTAA TCGGGTCGAG GCTCTGTTTC GGTTTGCCGG ATTTCGCGTA TGCTAAACGT
AACCTAGATT AGCCCAGCTC CGAGACAAAG CCAAACGGCC TAAAGCGCAT ACGATTTGCA

        760          770          780          790          800          810
         *            *            *            *            *            *
GCACACGCCA CAAACTAATT TAAGCTCCAA TTTAGATTAA ATAACAAATT ATCGTCGCTC
CGTGTGCGGT GTTTGATTAA ATTCGAGGTT AAATCTAATT TATTGTTTAA TAGCAGCGAG

        820          830          840          850          860          870
         *            *            *            *            *            *
TATTGTAGAT TTATTGTAAT AAAAGTGCAC TATTGATTTC ACATTCAAAA AAAAAAAAAA
ATAACATCTA AATAACATTA TTTTCACGTG ATAACTAAAG TGTAAGTTTT TTTTTTTTTT

AA
TT
```

Sequence Range: 1 to 739 (SEQ ID NO 23-24)

```
          10          20          30          40          50
           *           *           *           *           *
CCCCCCCCCC CCCCCCCCGA TTTTTTTTAA ATG GAC GAA ATC CTC TTT CCC ACG
GGGGGGGGGG GGGGGGGGCT AAAAAAAATT TAC CTG CTT TAG GAG AAA GGG TGC
                                  M   D   E   I   L   F   P   T>
                                    TRANSLATION OF 911G FULL    >

      60          70          80          90          100
       *           *           *           *           *
CAG CAA AAG AGC AAC TCC CTA AGC GAC GGC GAC GAT GTC GAC CTG AAA
GTC GTT TTC TCG TTG AGG GAT TCG CTG CCG CTG CTA CAG CTG GAC TTT
 Q   Q   K   S   N   S   L   S   D   G   D   D   V   D   L   K>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

       110         120         130         140         150
        *           *           *           *           *
TTC TTC CAG TCG GGC CTC CGG GGG AGG CGA AAG GAC AGC GAC ACC TCG
AAG AAG GTC AGC CCG GAG GCC CCC TCC GCT TTC CTG TCG CTG TGG AGC
 F   F   Q   S   G   L   R   G   R   R   K   D   S   D   T   S>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

          160         170         180         190
           *           *           *           *
GAT CCG GGA AAC GAT GCG GAT CGT GAT GGC AAA GAT GCG GAT GGG GAC
CTA GGC CCT TTG CTA CGC CTA GCA CTA CCG TTT CTA CGC CTA CCC CTG
 D   P   G   N   D   A   D   R   D   G   K   D   A   D   G   D>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

200         210         220         230         240
 *           *           *           *           *
AAC GAC AAC AAG AAC ACG GAC GGA GAT GGT GAC TCT GGC GAG CCG GCG
TTG CTG TTG TTC TTG TGC CTG CCT CTA CCA CTG AGA CCG CTC GGC CGC
 N   D   N   K   N   T   D   G   D   G   D   S   G   E   P   A>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

  250         260         270         280         290
   *           *           *           *           *
CAC AAA AAG CTC AAA ACC AAG AAG GAA CTG GAG GAG GAG GAG CGC GAA
GTG TTT TTC GAG TTT TGG TTC TTC CTT GAC CTC CTC CTC CTC GCG CTT
 H   K   K   L   K   T   K   K   E   L   E   E   E   E   R   E>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

    300         310         320         330         340
     *           *           *           *           *
CGA ATG CAG GTT CTC GTT TCC AAC TTT ACT GAA GAA CAG CTG GAT CGC
GCT TAC GTC CAA GAG CAA AGG TTG AAA TGA CTT CTT GTC GAC CTA GCG
 R   M   Q   V   L   V   S   N   F   T   E   E   Q   L   D   R>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

       350         360         370         380         390
        *           *           *           *           *
TAC GAA ATG TAT CGT CGC TCA GCC TTT CCC AAG GCC GCC GTC AAG CGT
ATG CTT TAC ATA GCA GCG AGT CGG AAA GGG TTC CGG CGG CAG TTC GCA
 Y   E   M   Y   R   R   S   A   F   P   K   A   A   V   K   R>
   a   a   a  TRANSLATION OF 911G FULL 5/20 [A]    a   a   a  >

          400         410         420         430
           *           *           *           *
CTA ATG CAA ACT ATC ACC GGC TGT TCC GTG TCC CAA AAT GTT GTG ATA
GAT TAC GTT TGA TAG TGG CCG ACA AGG CAC AGG GTT TTA CAA CAC TAT
```


FIG 9R

```
  L   M   Q   T   I   T   G   C   S   V   S   Q   N   V   V   I>
    a   a   a  TRANSLATION OF 911G FULL 5/20 (A)    a   a   a  >

440           450           460           470           480
 *             *             *             *             *
GCC ATG TCC GGC ATT GCG AAG GTC TTC GTC GGC GAG GTT GTG GAG GAA
CGG TAC AGG CCG TAA CGC TTC CAG AAG CAG CCG CTC CAA CAC CTC CTT
 A   M   S   G   I   A   K   V   F   V   G   E   V   V   E   E>
   a   a   a  TRANSLATION OF 911G FULL 5/20 (A)    a   a   a  >

  490           500           510           520           530
   *             *             *             *             *
GCC CTC GAC GTG ATG GAG GCC CAA GGT GAA TCC GGT GCC CTG CAG CCC
CGG GAG CTG CAC TAC CTC CGG GTT CCA CTT AGG CCA CGG GAC GTC GGG
 A   L   D   V   M   E   A   Q   G   E   S   G   A   L   Q   P>
   a   a   a  TRANSLATION OF 911G FULL 5/20 (A)    a   a   a  >

    540           550           560           570           580
     *             *             *             *             *
AAA TTC ATA CGA GAG GCA GTG CGA CGA CTG AGG ACC AAG GAT CGG ATG
TTT AAG TAT GCT CTC CGT CAC GCT GCT GAC TCC TGG TTC CTA GCC TAC
 K   F   I   R   E   A   V   R   R   L   R   T   K   D   R   M>
   a   a   a  TRANSLATION OF 911G FULL 5/20 (A)    a   a   a  >

      590           600           610           620           630
       *             *             *             *             *
CCC ATA GGC AGA TAC CAG CAG CCC TAT TTC AGA CTG AAC TAG C GAGTC
GGG TAT CCG TCT ATG GTC GTC GGG ATA AAG TCT GAC TTG ATC G CTCAG
 P   I   G   R   Y   Q   Q   P   Y   F   R   L   N   *   X>
   a   a  TRANSLATION OF 911G FULL 5/20 (A)    a   a   a  >

      640           650           660           670           680           690
       *             *             *             *             *             *
GAGACATTAA GAAATATAGT TTGTAAATCT GTTAGTGAAT ATAAAAATAC ATAAACAAGT
CTCTGTAATT CTTTATATCA AACATTTAGA CAATCACTTA TATTTTTATG TATTTGTTCA

      700           710           720           730
       *             *             *             *
AAAAAGTAAA TAAATATAAA GATTTTTTCA AGAAAAAAAA AAAAAAAAG
TTTTTCATTT ATTTATATTT CTAAAAAAGT TCTTTTTTTT TTTTTTTTC
```

```
          10          20          30          40
           *           *           *           *
ACC ATG TTG CTT CCG AAC ATC CTG CTC ACC GGT ACA CCA GGG GTT GGA
TGG TAC AAC GAA GGC TTG TAG GAC GAG TGG CCA TGT GGT CCC CAA CCT

50          60          70          80          90
 *           *           *           *           *
AAA ACC ACA CTA GGC AAA GAA CTT GCG TCA AAA TCA GGA CTG AAA TAC
TTT TGG TGT GAT CCG TTT CTT GAA CGC AGT TTT AGT CCT GAC TTT ATG

  100         110         120         130         140
   *           *           *           *           *
ATT AAT CTG GGT GAT TTA GCT CGA GAA GTC TGA TCA TCG GAT ATC ATG
TAA TTA GAC CCA CTA AAT CGA GCT CTT CAG ACT AGT AGC CTA TAG TAC
                                                    M>

    150         160         170         180         190
     *           *           *           *           *
GAG TCT GGC AAG ACG GCT TCT CCC AAG AGC ATG CCG AAA GAT GCA CAC
CTC AGA CCG TTC TGC CGA AGA GGG TTC TCG TAC GGC TTT CTA CGT GTG
 E   S   G   K   T   A   S   P   K   S   M   P   K   D   A   Q>

      200         210         220         230         240
       *           *           *           *           *
ATG ATG GCA CAA ATC CTG AAG GAT ATG GGG ATT ACA GAA TAT GAG CCA
TAC TAC CGT GTT TAG GAC TTC CTA TAC CCC TAA TGT CTT ATA CTC GGT
 M   M   A   Q   I   L   K   D   M   G   I   T   E   Y   E   P>

          250         260         270         280
           *           *           *           *
AGA GTT ATA AAT CAG ATG TTG GAC TTT GCC TTC CGA TAT GTG ACC ACA
TCT CAA TAT TTA GTC TAC AAC CTG AAA CGG AAG GCT ATA CAC TGG TGT
 R   V   I   N   Q   M   L   E   F   A   F   R   Y   V   T   T>

290         300         310         320         330
 *           *           *           *           *
ATT CTA GAT GAT GCA AAA ATT TAT TCA AGC CAT GCT AAG AAA GCT ACT
TAA GAT CTA CTA CGT TTT TAA ATA AGT TCG GTA CGA TTC TTT CGA TGA
 I   L   D   D   A   K   I   Y   S   S   H   A   K   K   A   T>

  340         350         360         370         380
   *           *           *           *           *
GTT GAT GCA GAT GAT GTG CGA TTG GCA ATC CAG TGC CGC GCT GAT CAG
CAA CTA CGT CTA CTA CAC GCT AAC CGT TAG GTC ACG GCG CGA CTA GTC
 V   D   A   D   D   V   R   L   A   I   Q   C   R   A   D   Q>

    390         400         410         420         430
     *           *           *           *           *
TCT TTT ACC TCT CCT CCC CCA AGA GAT TTT TTA TTA GAT ATT GCA AGG
AGA AAA TGG AGA GGA GGG GGT TCT CTA AAA AAT AAT CTA TAA CGT TCC
 S   F   T   S   P   P   P   R   D   F   L   L   D   I   A   R>

      440         450         460         470         480
       *           *           *           *           *
CAA AGA AAT CAA ACC CCT TTG CCA TTG ATC AAG CCA TAT TCA GGT CCT
GTT TCT TTA GTT TGG GGA AAC GGT AAC TAG TTC GGT ATA AGT CCA GGA
 Q   R   N   Q   T   P   L   P   L   I   K   P   Y   S   G   P>

          490         500         510         520
           *           *           *           *
```

```
AGC TTG CCA CCT GAT AGA TAC TGC TTA ACA GCT CCA AAC TAT AGG CTG
TCC AAC GGT GGA CTA TCT ATG ACG AAT TGT CGA GGT TTG ATA TCC GAC
 R   L   P   P   D   R   Y   C   L   T   A   P   N   Y   R   L>

530         540         550         560         570
 *           *           *           *           *
AAA TCT TTA CAG AAA AAG GCA TCA ACT TCT GCG GGA AGA ATA ACA GTC
TTT AGA AAT GTC TTT TTC CGT AGT TGA AGA CGC CCT TCT TAT TGT CAG
 K   S   L   Q   K   K   A   S   T   S   A   G   R   I   T   V>

  580         590         600         610         620
   *           *           *           *           *
CCG CGG TTA AGT GTT GGT TCA GTT ACT AGC AGA CCA AGT ACT CCC ACA
GGC GCC AAT TCA CAA CCA AGT CAA TGA TCG TCT GGT TCA TGA GGG TGT
 P   R   L   S   V   G   S   V   T   S   R   P   S   T   P   T>

    630         640         650         660         670
     *           *           *           *           *
CTA GGC ACA CCA ACC CCA CAG ACC ATG TCT GTT TCA ACT AAA GTA GGG
GAT CCG TGT GGT TGG GGT GTC TGG TAC AGA CAA AGT TGA TTT CAT CCC
 L   G   T   P   T   P   Q   T   M   S   V   S   T   K   V   G>

        680         690         700         710         720
         *           *           *           *           *
ACT CCC ATG TCC CTC ACA GGT CAA AGG TTT ACA GTA CAG ATG CCT ACT
TGA GGG TAC AGG GAG TGT CCA GTT TCC AAA TGT CAT GTC TAC GGA TGA
 T   P   M   S   L   T   G   Q   R   F   T   V   Q   M   P   T>

          730         740         750         760
           *           *           *           *
TCT CAG TCT CCA GCT GTA AAA GCT TCA ATT CCT GCA ACC TCA GCA GTT
AGA GTC AGA GGT CGA CAT TTT CGA AGT TAA GGA CGT TGG AGT CGT CAA
 S   Q   S   P   A   V   K   A   S   I   P   A   T   S   A   V>

770         780         790         800         810
 *           *           *           *           *
CAG AAT GTT CTG ATT AAT CCA TCA TTA ATC GGG TCC AAA AAC ATT CTT
GTC TTA CAA GAC TAA TTA GGT AGT AAT TAG CCC AGG TTT TTG TAA GAA
 Q   N   V   L   I   N   P   S   L   I   G   S   K   N   I   L>

  820         830         840         850         860
   *           *           *           *           *
ATT ACC ACT AAT ATG ATG TCA TCA CAA AAT ACT GCC AAT GAA TCA TCA
TAA TGG TGA TTA TAC TAC AGT AGT GTT TTA TGA CGG TTA CTT AGT AGT
 I   T   T   N   M   M   S   S   Q   N   T   A   N   E   S   S>

    870         880         890         900         910
     *           *           *           *           *
AAT GCA TTG AAA AGA AAA CGT GAA GAT GAT GAT GAT GAC GAT GAT GAT
TTA CGT AAC TTT TCT TTT GCA CTT CTA CTA CTA CTA CTG CTA CTA CTA
 N   A   L   K   R   K   R   E   D   D   D   D   D   D   D   D>

        920         930         940         950         960
         *           *           *           *           *
GAT GAT GAC TAT GAT AAT CTG TAA TCT AGC CTT GCT GAA TGT AAC ATG
CTA CTA CTG ATA CTA TTA GAC ATT AGA TCG GAA CGA CTT ACA TTG TAC
 D   D   D   Y   D   N   L>

          970         980         990         1000
           *           *           *           *
TAT ACT TGG TCT TGA ATT CAT TGT ACT GAT ATT AAA CAT GCA TGC TGG
ATA TGA ACC AGA ACT TAA GTA ACA TGA CTA TAA TTT GTA CGT ACG ACC
```

```
1010        1020        1030        1040        1050
 *           *           *           *           *
ATC TTT TCA AGT TGT GTT TTA GAA AAC TAA TAA TAA TGA GTA AAC ACA
TAC AAA AGT TCA ACA CAA AAT CTT TTG ATT ATT ATT ACT CAT TTG TGT

1060        1070        1080        1090        1100
 *           *           *           *           *
GTT ACC ATA CTT TTC AAT TGA AAT GAA GGT TTT TCA TCA GCC TTA AAA
CAA TGG TAT GAA AAG TTA ACT TTA CTT CCA AAA AGT ACT CGG AAT TTT

1110        1120        1130        1140        1150
 *           *           *           *           *
GTG TAA GAA AAA TAA AGT TGT CAT TCA TTC GAT AAA AAA AAA AAA A
CAC ATT CTT TTT ATT TCA ACA GTA AGT AAG CTA TTT TTT TTT TTT T
```

hTAFII30α peptides: (SEQ ID NO 27)

1. DVQLHLERQ_NM_IPGFGSEEI_PYK
2. KKLQDLVREVDPNEQLDEDV_EMLLQIADD
3. LQDLVREVDPN

hTAFII30β peptides: (SEQ ID NO 28)

1. VFVGEVVEEALDVEEKP
2. HMREAVRRLK
3. MQILVSSFEEEQLN_YEMYN_K_AYGQ

hTAF I 48 -> Genes

DNA sequence 1578 b.p. ATTCCAAGCTAA ... GTCTGTTTTCTT linear

Read from Bionet/Intelligenetics file "48 prot"

```
   1 ATTCCAAGCTAAATTTAGGCGGGT ATG AGT GAT TTC AGT GAA GAA TTA AAA GGG CCT GTG ACA GAT  66
   1                          M   S   D   F   S   E   E   L   K   G   P   V   T   D   14

  67 GAT GAA GAA GTG GAA ACA TCT GTG CTC AGT GGT GCA GGA ATG CAT TTT CCT TGG CTT CAA  126
  15 D   E   E   V   E   T   S   V   L   S   G   A   G   M   H   F   P   W   L   Q   34

 127 ACA TAC GTA GAA ACT GTG GCC ATT GGA GGG AAA AGG AGG AAG GAT TTT GCT CAG ACA ACA  186
  35 T   Y   V   E   T   V   A   I   G   G   K   R   R   K   D   F   A   Q   T   T   54

 187 AGT GCT TGT TTA AGT TTT ATC CAA GAA GCT CTG CTG AAG CAC CAA TGG CAG CAA GCT GCA  246
  55 S   A   C   L   S   F   I   Q   E   A   L   L   K   H   Q   W   Q   Q   A   A   74

 247 GAA TAC ATG TAC AGT TAT TTT CAG ACC TTG GAA GAT TCA GAT AGC TAC AAA AGG CAG GCT  306
  75 E   Y   M   Y   S   Y   F   Q   T   L   E   D   S   D   S   Y   K   R   Q   A   94

 307 GCA CCT GAG ATT ATT TGG AAG CTC GGA AGT GAA ATT CTA TTT TAT CAT CCC AAA AGC AAC  366
  95 A   P   E   I   I   W   K   L   G   S   E   I   L   F   Y   H   P   K   S   N  114

 367 ATG GAG AGT TTC AAT ACT TTT GCT AAC CGG ATG AAA AAT ATT GGC GTC ATG AAT TAT TTA  426
 115 M   E   S   F   N   T   F   A   N   R   M   K   N   I   G   V   M   N   Y   L  134

 427 AAG ATC TCC TTA CAA CAT GCA TTA TAC CTT CTG CAT CAT GGA ATG CTT AAA GAT GCT AAG  486
 135 K   I   S   L   Q   H   A   L   Y   L   L   H   H   G   M   L   K   D   A   K  154

 487 AGA AAT CTG AGT GAG GCA GAG ACA TGG AGA CAT GGT GAA AAT ACG TCT TCC CGG GAA ATA  546
 155 R   N   L   S   E   A   E   T   W   R   H   G   E   N   T   S   S   R   E   I  174

 547 TTA ATC AAC CTT ATT CAG GCC TAT AAA GGG CTT TTA CAG TAT TAT ACC TGG TCT GAA AAG  606
 175 L   I   N   L   I   Q   A   Y   K   G   L   L   Q   Y   Y   T   W   S   E   K  194

 607 AAG ATG GAA TTG TCA AAG CTT GAT AAG GAT GAT TAT GCT TAC AAT GCA GTA GCC CAG GAT  666
 195 K   M   E   L   S   K   L   D   K   D   D   Y   A   Y   N   A   V   A   Q   D  214

 667 GTG TTC AAC CAC AGC TGG AAG ACA TCT GCA AAT ATT TCT GCA TTG ATT AAA ATT CCT GGA  726
 215 V   F   N   H   S   W   K   T   S   A   N   I   S   A   L   I   K   I   P   G  234

 727 GTT TGG GAC CCT TTT GTG AAG AGT TAT GTA GAA ATG CTG GAA TTC TAT GGG GAT CGA GAT  786
 235 V   W   D   P   F   V   K   S   Y   V   E   M   L   E   F   Y   G   D   R   D  254

 787 GGA GCC CAA GAG GTA CTC ACC AAT TAT GCA TAT GAT GAA AAG TTT CCA TCA AAT CCA AAT  846
 255 G   A   Q   E   V   L   T   N   Y   A   Y   D   E   K   F   P   S   N   P   N  274

 847 GCC CAT ATC TAC TTA TAC AAC TTT CTA AAG AGA CAG AAG GCA CCA AGA TCA AAA TTG ATA  906
 275 A   H   I   Y   L   Y   N   F   L   K   R   Q   K   A   P   R   S   K   L   I  294

 907 AGT GTG CTT AAG ATT TTG TAT CAG ATT GTA CCA TCT CAT AAA TTG ATG TTG GAA TTC CAT  966
 295 S   V   L   K   I   L   Y   Q   I   V   P   S   H   K   L   M   L   E   F   H  314

 967 ACA TTA CTT AGA AAA TCA GAA AAA GAA GAA CAC CGT AAA CTG GGG TTG GAG GTA TTA TTT 1026
 315 T   L   L   R   K   S   E   K   E   E   H   R   K   L   G   L   E   V   L   F  334

1027 GGA GTC TTA GAT TTT GCC GGA TGC ACT AAG AAT ATA ACT GCT TGG AAA TAC TTG GCA AAA 1086
 335 G   V   L   D   F   A   G   C   T   K   N   I   T   A   W   K   Y   L   A   K  354

1087 TAT CTG AAA AAT ATC TTA ATG GGA AAC CAC CTT GCG TGG GTT CAA GAA GAG TGG AAC TCC 1146
 355 Y   L   K   N   I   L   M   G   N   H   L   A   W   V   Q   E   E   W   N   S  374

1147 AGG AAA AAC TGG TGG CCA GGG TTT CAT TTC AGC TAC TTT TGG GCA AAA AGT GAT TGG AAG 1206
 375 R   K   N   W   W   P   G   F   H   F   S   Y   F   W   A   K   S   D   W   K  394

1207 GAA GAT ACA GCT TTG GCC TGT GAG AAA GCT TTT GTG GCT GGT TTA CTG TTA GGA AAA GGT 1266
 395 E   D   T   A   L   A   C   E   K   A   F   V   A   G   L   L   L   G   K   G  414

1267 TGT AGA TAT TTC CGG TAT ATT TTA AAG CAA GAT CAC CAA ATC TTA GGG AAG AAA ATT AAG 1326
 415 C   R   Y   F   R   Y   I   L   K   Q   D   H   Q   I   L   G   K   K   I   K  434

1327 CGG ATG AAG AGA TCT GTG AAA AAA TAC AGT ATT GTA AAT CCA AGA CTC TGA TACTGAATTTTA 1389
 435 R   M   K   R   S   V   K   K   Y   S   I   V   N   P   R   L   *               451
```

FIG 11

hTAF$_I$110 cDNA and deduced amino acid sequence

Dr. R. Tjian laboratory, Department of Molecular and Cell Biology,
University of California, Berkeley.

```
        10        20        30        40        50
         *         *         *         *         *
GCTCGAGTGCCAAAGCTGGGGTTCTACTTGAGATTTCCCTCGTGGTGCCA

        60        70        80        90       100
         *         *         *         *         *
GGGTCCGGCGAGCATCACGCCGAGGCCCATTTTCCAGACGACCACGACGA

       110       120       130       140       150
         *         *         *         *         *
GGCCGGGGTCACGAACTCTGGCGCCCCTTACCAGCTTCCAGTCTCTCGAG

       160       170       180       190       200
         *         *         *         *         *
GTGGCCAGTGTGGTGCTTGGTCCTTGTTTCCAGGATGGACTTCCCCAGCT
                                    M  D  F  P  S>

       210       220       230       240       250
         *         *         *         *         *
CCCTCCGCCCTGCGTTGTTTCTGACCGGCCCCCTTGGTCTGAGCGACGTC
S  L  R  P  A  L  F  L  T  G  P  L  G  L  S  D  V>

       260       270       280       290       300
         *         *         *         *         *
CCTGACCTCTCTTTCATGTGCAGCTGGCGAGACGCACTGACTCTGCCAGA
 P  D  L  S  F  M  C  S  W  R  D  A  L  T  L  P  E>

       310       320       330       340       350
         *         *         *         *         *
GGCCCAGCCCCAGAACTCAGAGAATGGGGCACTGCATGTGACCAAGGACC
  A  Q  P  Q  N  S  E  N  G  A  L  H  V  T  K  D>

       360       370       380       390       400
         *         *         *         *         *
TGCTGTGGGAGCCGGCAACCCCTGGGCCTCTCCCCATGCTGCCTCCCCTC
L  L  W  E  P  A  T  P  G  P  L  P  M  L  P  P  L>

       410       420       430       440       450
         *         *         *         *         *
ATCGATCCCTGGGACCCTGGCCTGACTGCCCGGGACCTGCTTTTCCGCGG
 I  D  P  W  D  P  G  L  T  A  R  D  L  L  F  R  G>

       460       470       480       490       500
         *         *         *         *         *
AGGGTACCGGTATCGGAAGCGGCCCCGAGTCGTGCTGGATGTGACTGAGC
  G  Y  R  Y  R  K  R  P  R  V  V  L  D  V  T  E>
```


Fig 12

```
       510        520        530        540        550
        *          *          *          *          *
AGATCAGCCGGTTCCTCTTGGATCATGGAGACGTAGCCTTTGCGCCCCTG
Q  I  S  R  F  L  L  D  H  G  D  V  A  F  A  P  L>

       560        570        580        590        600
        *          *          *          *          *
GGGAAGCTGATGCTGGAGAATTTCAAGCTGGAGGGAGCGGGGAGCCGCAC
 G  K  L  M  L  E  N  F  K  L  E  G  A  G  S  R  T>

       610        620        630        640        650
        *          *          *          *          *
TAAGAAGAAGACAGTGGTCAGTGTGAAGAAGCTGCTCCAGGACCTCGGTG
 K  K  K  T  V  V  S  V  K  K  L  L  Q  D  L  G>

       660        670        680        690        700
        *          *          *          *          *
GACACCAGCCCTGGGGGTGTCCCTGGGCTTACCTCAGCAACCGACAGCGC
G  H  Q  P  W  G  C  P  W  A  Y  L  S  N  R  Q  R>

       710        720        730        740        750
        *          *          *          *          *
CGCTTCTCTATCCTCGGGGGCCCCATCCTGGGCACGTCGGTGGCGAGCCA
 R  F  S  I  L  G  G  P  I  L  G  T  S  V  A  S  H>

       760        770        780        790        800
        *          *          *          *          *
CTTGGCAGAGCTGCTGCACGAGGAGCTGGTGCTGCGGTGGGAGCAGCTGC
  L  A  E  L  L  H  E  E  L  V  L  R  W  E  Q  L>

       810        820        830        840        850
        *          *          *          *          *
TTCTGGATGAGGCCTGCACTGGGGGCGCGCTGGCCTGGGTTCCTGGAAGG
L  L  D  E  A  C  T  G  G  A  L  A  W  V  P  G  R>

       860        870        880        890        900
        *          *          *          *          *
ACACCCCAGTTCGGGCAGCTGGTCTACCCTGCTGGAGGCGCCCAGGACAG
 T  P  Q  F  G  Q  L  V  Y  P  A  G  G  A  Q  D  R>

       910        920        930        940        950
        *          *          *          *          *
GCTGCATTTCCAAGAGGTCGTTCTGACCCCAGGTGACAATCCCCAATTCC
  L  H  F  Q  E  V  V  L  T  P  G  D  N  P  Q  F>

       960        970        980        990       1000
        *          *          *          *          *
TTGGGAAACCTGGACGCATCCAGCTCCAGGGACCTGTCCGGCAAGTGGTG
L  G  K  P  G  R  I  Q  L  Q  G  P  V  R  Q  V  V>

      1010       1020       1030       1040       1050
        *          *          *          *          *
ACATGCACCGTCCAGGGAGAAAGTAAGGCCCTTATATACACTTTCCTCCC
 T  C  T  V  Q  G  E  S  K  A  L  I  Y  T  F  L  P>
```

```
      1060       1070       1080       1090       1100
        *          *          *          *          *
TCACTGGCTGACCTGCTACCTGACCCCTGGCCCTTTCCATCCCTCCTCAG
  H  W  L  T  C  Y  L  T  P  G  P  F  H  P  S  S>

      1110       1120       1130       1140       1150
        *          *          *          *          *
CTCTGCTGGCCGTCCGCTCTGACTACCACTGTGCCGTGTGGAAGTTTGGT
A  L  L  A  V  R  S  D  Y  H  C  A  V  W  K  F  G>

      1160       1170       1180       1190       1200
        *          *          *          *          *
AAACAGTGGCAGCCAACCCTTCTGCAGGCGATGCAGGTGGAGAAAGGGGC
 K  Q  W  Q  P  T  L  L  Q  A  M  Q  V  E  K  G  A>

      1210       1220       1230       1240       1250
        *          *          *          *          *
CACGGGGATCAGCCTCAGCCCTCACCTGCCCGGGGAGCTGGCCATCTGCA
  T  G  I  S  L  S  P  H  L  P  G  E  L  A  I  C>

      1260       1270       1280       1290       1300
        *          *          *          *          *
GCCGCTCGGGAGCCGTCTGCCTGTGGAGCCCTGAGGATGGGCTGCGGCAA
S  R  S  G  A  V  C  L  W  S  P  E  D  G  L  R  Q>

      1310       1320       1330       1340       1350
        *          *          *          *          *
ATCTACAGGGACCCTGAGACCCTCGTGTTCCGGGACTCCTCTTCGTGGCG
 I  Y  R  D  P  E  T  L  V  F  R  D  S  S  S  W  R>

      1360       1370       1380       1390       1400
        *          *          *          *          *
TTGGGCAGACTTCACTGCGCACCCTCGGGTGCTGACCGTGGGTGACCGCA
  W  A  D  F  T  A  H  P  R  V  L  T  V  G  D  R>

      1410       1420       1430       1440       1450
        *          *          *          *          *
CCGGAGTGAAGATGCTGGACACTCAGGGCCCGCCGGGCTGTGGTCTGTTG
T  G  V  K  M  L  D  T  Q  G  P  P  G  C  G  L  L>

      1460       1470       1480       1490       1500
        *          *          *          *          *
CTTTTTCGTTTGGGGGCAGAGGCTTCGTGCCAGAAAGGGGAACGTGTCCT
 L  F  R  L  G  A  E  A  S  C  Q  K  G  E  R  V  L>

      1510       1520       1530       1540       1550
        *          *          *          *          *
GCTTACCCAGTACCTGGGGCACTCCAGCCCCAAATGCCTCCCCCCTACTC
  L  T  Q  Y  L  G  H  S  S  P  K  C  L  P  P  T>

      1560       1570       1580       1590       1600
        *          *          *          *          *
TTCATCTCGTCTGTACCCAGTTCTCTCTCTACCTAGTGGACGAGCGCCTT
L  H  L  V  C  T  Q  F  S  L  Y  L  V  D  E  R  L>
```

```
      1610       1620       1630       1640       1650
        *          *          *          *          *
CCCCTGGTGCCGATGCTGAAGTGGAACCATGGCCTCCCCTCCCCGCTCCT
 P  L  V  P  M  L  K  W  N  H  G  L  P  S  P  L  L>

      1660       1670       1680       1690       1700
        *          *          *          *          *
GCTGGCCCGACTGCTGCCTCCGCCCCGGCCCAGCTGCGTGCAGCCCCTGC
  L  A  R  L  L  P  P  P  R  P  S  C  V  Q  P  L>

      1710       1720       1730       1740       1750
        *          *          *          *          *
TCCTCGGAGGCCAGGGTGGGCAGCTGCAGCTGCTGCACCTGGCAGGAGAA
L  L  G  G  Q  G  G  Q  L  Q  L  L  H  L  A  G  E>

      1760       1770       1780       1790       1800
        *          *          *          *          *
GGGGCGTCGGTGCCCCGCCTGGCAGGCCCCCCCCAGTCTCTTCCTTCCAG
 G  A  S  V  P  R  L  A  G  P  P  Q  S  L  P  S  R>

      1810       1820       1830       1840       1850
        *          *          *          *          *
GATCGACTCCCTCCCTGCATTTCCTCTGCTGGAGCCTAAGATCCAGTGGC
  I  D  S  L  P  A  F  P  L  L  E  P  K  I  Q  W>

      1860       1870       1880       1890       1900
        *          *          *          *          *
GGCTGCAGGAGCGCCTGAAAGCACCGACCATAGGTCTGGCTGCCGTCGTC
R  L  Q  E  R  L  K  A  P  T  I  G  L  A  A  V  V>

      1910       1920       1930       1940       1950
        *          *          *          *          *
CCGCCCTTGCCCTCAGCGCCCACACCAGGCCTGGTGCTCTTCCAGCTCTC
 P  P  L  P  S  A  P  T  P  G  L  V  L  F  Q  L  S>

      1960       1970       1980       1990       2000
        *          *          *          *          *
GGCGGCGGGAGATGTCTTCTACCAGCAGCTCCGCCCCCAGGTGGACTCCA
  A  A  G  D  V  F  Y  Q  Q  L  R  P  Q  V  D  S>

      2010       2020       2030       2040       2050
        *          *          *          *          *
GCCTCCGCAGAGATGCTGGGCCTCCTGGCGACACCCAACCTGACTGCCAT
S  L  R  R  D  A  G  P  P  G  D  T  Q  P  D  C  H>

      2060       2070       2080       2090       2100
        *          *          *          *          *
GCCCCCACAGCTTCCTGGACCTCCCAGGACACTGCCGGCTGCAGCCAGTG
 A  P  T  A  S  W  T  S  Q  D  T  A  G  C  S  Q  W>

      2110       2120       2130       2140       2150
        *          *          *          *          *
GCTGAAGGCCCTGCTAAAAGTGCCCCTGGCTCCTCCTGTGTGGACAGCAC
  L  K  A  L  L  K  V  P  L  A  P  P  V  W  T  A>
```

```
      2160       2170       2180       2190       2200
        *          *          *          *          *
CCACCTTCACCCACCGCCAGATGCTGGGCAGCACAGAGCTGCGGAGGGAG
P  T  F  T  H  R  Q  M  L  G  S  T  E  L  R  R  E>

      2210       2220       2230       2240       2250
        *          *          *          *          *
GAAGAGGAAGGGCAGCGGCTGGGTGTGCTCCGCAAGGCCATGGCCCGAGG
 E  E  E  G  Q  R  L  G  V  L  R  K  A  M  A  R  G>

      2260       2270       2280       2290       2300
        *          *          *          *          *
GCAGCTCCTGCTGCAGAGAGACCTGGGCTCCCTCCCTGCGGCAGAGCCAC
  Q  L  L  L  Q  R  D  L  G  S  L  P  A  A  E  P>

      2310       2320       2330       2340       2350
        *          *          *          *          *
CCCCTGCACCCGAGTCAGGCCTAGAGGACAAGCTCAGTGAGCGCCTGGGG
P  P  A  P  E  S  G  L  E  D  K  L  S  E  R  L  G>

      2360       2370       2380       2390       2400
        *          *          *          *          *
GAAGCCTGGGCAGGCCGAGGGGCTGCCTGGTGGGAGAGGCAGCAGGGCAG
 E  A  W  A  G  R  G  A  A  W  W  E  R  Q  Q  G  R>

      2410       2420       2430       2440       2450
        *          *          *          *          *
GACCTCGGAGCCCGGGAGACAGACCAGGCGGCCCAAGCGCCGGACCCAGC
  T  S  E  P  G  R  Q  T  R  R  P  K  R  R  T  Q>

      2460       2470       2480       2490       2500
        *          *          *          *          *
TGTCCAGCAGCTTTTCGCTCAGTGGCCATGTGGATCCGTCAGAGGACACC
L  S  S  S  F  S  L  S  G  H  V  D  P  S  E  D  T>

      2510       2520       2530       2540       2550
        *          *          *          *          *
AGCTCCCCTCATAGCCCTGAGTGGCCACCTGCTGATGCTCTGCCCCTGCC
 S  S  P  H  S  P  E  W  P  P  A  D  A  L  P  L  P>

      2560       2570       2580       2590       2600
        *          *          *          *          *
CCCCACGACCCCGCCCTCCCAGGAGTTGACTCCGGATGCATGCGCCCAGG
  P  T  T  P  P  S  Q  E  L  T  P  D  A  C  A  Q>

      2610       2620       2630       2640       2650
        *          *          *          *          *
GCGTCCCATCAGAGCAGCGGCAGATGCTCCGTGACTACATGGCCAAGCTA
G  V  P  S  E  Q  R  Q  M  L  R  D  Y  M  A  K  L>

      2660       2670       2680       2690       2700
        *          *          *          *          *
CCACCCCAGAGGGACACCCCAGGCTGTGCCACCACACCTCCCCACTCCCA
 P  P  Q  R  D  T  P  G  C  A  T  T  P  P  H  S  Q>
```

```
      2710       2720       2730       2740       2750
        *          *          *          *          *
GGCCTCCAGCGTCCGGGCCACTCGCTCCCAGCAGCACACACCCGTCCTCT
  A  S  S  V  R  A  T  R  S  Q  Q  H  T  P  V  L>

      2760       2770       2780       2790       2800
        *          *          *          *          *
CTAGCTCTCAGCCCCTCCGGAAGAAGCCTCGAATGGGCTTCTGAGGACAC
S  S  S  Q  P  L  R  K  K  P  R  M  G  F>

      2810       2820       2830       2840       2850
        *          *          *          *          *
AAGGTGGGCTGCCCTCAAGCCCCAGAGAGCCCCTCATCCTTCCTCTGGGA

      2860       2870       2880       2890       2900
        *          *          *          *          *
CCAGATGTGCCTTCCACAGTTGAAACTTGAGAAGCAGAGCTCGCCACCTT

      2910       2920       2930       2940       2950
        *          *          *          *          *
CTGGAGGCCACTGTGATGATGAGCCAAGCAATTTGGAGCCAAGTTGAAGG

      2960       2970       2980       2990       3000
        *          *          *          *          *
GACAGGGCAACAAAATACAGTAGTAGTTTCTTTTGTATTTTGTATATTCG

      3010       3020       3030       3040       3050
        *          *          *          *          *
CCTGAAGATCATCCCGCAAGGCAGGCTGGAGGTGCCGGTGGGCCTGTGTT

      3060       3070       3080       3090       3100
        *          *          *          *          *
GCTGGGATTTTAGTCTGTGCTGGGAGGCAGGGCTCCGTGCGCCTCAGCTG

      3110       3120       3130       3140       3150
        *          *          *          *          *
TGGGGGCCTCAGGCAGGTCCCTCAGTTCTCACGCCTTCCTGTCCAGTGGA

      3160       3170       3180       3190       3200
        *          *          *          *          *
ATGGGGGCCAGGAGTGCTGGCTCCTCGTGTTTGGTGAGGGTGGAGTGAGG

      3210       3220       3230       3240       3250
        *          *          *          *          *
CCCCTGCAGAGCTGCTGATGAGGTGGGCACAGCGGCCGTTGGCAGCTGCT

      3260       3270       3280       3290       3300
        *          *          *          *          *
GTTGTGGGTTGCTTTGTCAATCTCTGCCCCGGTCTGATGTTTCCTACAGG

      3310       3320       3330       3340       3350
        *          *          *          *          *
GAGATGCCGTGGATCCAGGTTCAGGGACTAAATACACTTGGCAGCTGAAG
```

```
      3360       3370       3380       3390       3400
        *          *          *          *          *
ATGAATTGGAATGGTCACGTTTTTTAGGCTGGNACAGCGTCCCGCCACAG

      3410       3420       3430       3440       3450
        *          *          *          *          *
CTACTACCTGACACTGAGCTCATGCAGAGAGATGATGGCTGATGTTCCTT

      3460       3470       3480       3490       3500
        *          *          *          *          *
CTCCCTTGGGACATGGGTCTGGCACCTGTGGGCTGTCGATAGTGCCCTCT

      3510       3520       3530       3540       3550
        *          *          *          *          *
GAGCAGAGGGTCACGGTCATGTCAGTTTGGGGGAATTCTCTGTTGTGCCT

      3560       3570       3580       3590       3600
        *          *          *          *          *
CAGAGACTCCCCCCTTTCTTTCCTCCCTCCCCTTCTCATTTTGATGTCTA

      3610       3620       3630       3640       3650
        *          *          *          *          *
AAGCATCAAGTCCCTCTTCCTCAGAGTTTCTCTAGCTGCAGTGGAAGATT

      3660       3670       3680       3690       3700
        *          *          *          *          *
CTGTTTTCCTGTGGGGAAAATGCTCACTTGAGATTTTGCAGGGACCCGGG

      3710       3720       3730       3740       3750
        *          *          *          *          *
TCTGTCTGGTTTCTGATGACATAGTAAGAGAAAGGTCTTTTTTCAGGTTG

      3760       3770       3780       3790       3800
        *          *          *          *          *
GCTGGTGAAAGGAATTGCATGTGACTCACACAAACAGGAGCTAGCCCAAT

      3810       3820       3830       3840       3850
        *          *          *          *          *
CATACACTGACTCGCGTGGGTGTTTAAATGTTTATCATGCCTAAGGGAGA

      3860       3870       3880       3890       3900
        *          *          *          *          *
CATTTATAATTAAACCATTTATGCTACATAAAAAAAAAAAAAAAAAAAAA

AA
```

```
      3360       3370       3380       3390       3400
        *          *          *          *          *
ATGAATTGGAATGGTCACGTTTTTTAGGCTGGNACAGCGTCCCGCCACAG

      3410       3420       3430       3440       3450
        *          *          *          *          *
CTACTACCTGACACTGAGCTCATGCAGAGAGATGATGGCTGATGTTCCTT

      3460       3470       3480       3490       3500
        *          *          *          *          *
CTCCCTTGGGACATGGGTCTGGCACCTGTGGGCTGTCGATAGTGCCCTCT

      3510       3520       3530       3540       3550
        *          *          *          *          *
GAGCAGAGGGTCACGGTCATGTCAGTTTGGGGGAATTCTCTGTTGTGCCT

      3560       3570       3580       3590       3600
        *          *          *          *          *
CAGAGACTCCCCCCTTTCTTTCCTCCCTCCCCTTCTCATTTTGATGTCTA

      3610       3620       3630       3640       3650
        *          *          *          *          *
AAGCATCAAGTCCCTCTTCCTCAGAGTTTCTCTAGCTGCAGTGGAAGATT

      3660       3670       3680       3690       3700
        *          *          *          *          *
CTGTTTTCCTGTGGGGAAAATGCTCACTTGAGATTTTGCAGGGACCCGGG

      3710       3720       3730       3740       3750
        *          *          *          *          *
TCTGTCTGGTTTCTGATGACATAGTAAGAGAAAGGTCTTTTTTCAGGTTG

      3760       3770       3780       3790       3800
        *          *          *          *          *
GCTGGTGAAAGGAATTGCATGTGACTCACACAAACAGGAGCTAGCCCAAT

      3810       3820       3830       3840       3850
        *          *          *          *          *
CATACACTGACTCGCGTGGGTGTTTAAATGTTTATCATGCCTAAGGGAGA

      3860       3870       3880       3890       3900
        *          *          *          *          *
CATTTATAATTAAACCATTTATGCTACATAAAAAAAAAAAAAAAAAAAAA

AA
```

WHAT IS CLAIMED IS:

1. A composition comprising a substantially pure, biologically active portion of a TAF, wherein said TAF is other than CCG1.

2. A composition according to claim 1 wherein said portion is a substantially full-length TAF.

3. A composition according to claim 1 wherein said TAF is selected from the group consisting of dTAF250, dTAFII150, dTAFII110, dTAFII80, dTAFII60, dTAFII40 and dTAFII30.

4. A composition according to claim 1 wherein said TAF is selected from the group consisting of hTAFII250, hTAFII150, hTAFII130, hTAFII100, hTAFII70, hTAFII40 and hTAFII30.

5. A composition according to claim 1 wherein said TAF is selected from the group consisting of hTAFI110, hTAFI63 and hTAFI48.

6. A composition according to claim 1 wherein said TAF is selected from the group consisting of hTAFIII172, and hTAFIII25.

7. A composition comprising an isolated nucleic acid sequence encoding a portion of a TAF according to claim 1.

8. A composition according to claim 7 wherein said portion is a substantially full-length TAF.

9. A composition according to claim 7 wherein said TAF is selected from the group consisting of dTAF250, dTAFII150, dTAFII110, dTAFII80, dTAFII60, dTAFII40 and dTAFII30.

10. A composition according to claim 7 wherein said TAF is selected from the group consisting of hTAFII250, hTAFII150, hTAFII130, hTAFII100, hTAFII70, hTAFII40 and hTAFII30.

11. A composition according to claim 7 wherein said TAF is selected from the group consisting of hTAFII110, hTAFI63 and hTAFI48.

12. A composition according to claim 7 wherein said TAF is selected from the group consisting of hTAFIII172, and hTAFIII25.

13. An antibody that specifically binds a composition according to Claim 1.

14. A vector comprising a nucleic acid sequence according to claim 7 operably linked to a transcription regulatory element.

15. A cell comprising a nucleic acid sequence according to claim 7.

16. A process for the production of a TAF comprising culturing the cell of Claim 15 under conditions suitable for the expression of said TAF and recovering said TAF.

17. A composition comprising a recombinantly produced TAF.

18. A method of identifying an agent useful in the diagnosis or treatment of disease associated with transcription, said method comprising the steps of:

contacting an agent with at least a portion of a TAF according to claim 1; and,

determining whether said agent specifically binds said TAF.

19. A method of identifying an agent useful in the diagnosis or treatment of disease associated with transcription, said method comprising the steps of:

adding an agent to a mixture comprising at least a portion of a TAF according to claim 1:

comparing the association of mixture components before and after said adding step:

identifying an agent that alters the association of mixture components.

20. A method for treating disease, said method comprising:
identifying an agent according to the method of claim 18; and,
contacting an individual with said agent;
wherein said agent modulates transcription in said individual.

21. A method for treating disease, said method comprising:
identifying an agent according to the method of claim 19; and,
contacting an individual with said agent;
wherein said agent modulates transcription in said individual.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/01114

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 69.1, 172.3, 240.1, 320.1; 530/350, 388.1, 388.85; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Cell, Volume 66, issued 09 August 1991, Dynlacht et al, "Isolation of Coactivators Associated with the TATA-Binding Protein That Mediate Transcriptional Activation", pages 563-576, see entire document.	1-21
A	Nature, Volume 340, issued 1989, Fields et al, "A Novel Genetic System to Detect Protein-Protein Interactions", pages 245-246.	1-21
Y	International Journal of Pharmaceutics, Volume 72, issued 1991, Gambari et al, "TAPP (tetra-*p*-amidinophenoxyneopentane) Inhibits the Binding of Nuclear Factors to Target DNA Sequences", pages 251-258, see entire document.	18, 19

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
27 MAY 1994	JUN 22 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DOUGLAS GURIAN-SHERMAN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/01114

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the National Academy of Sciences, USA, Volume 89, issued December 1992, Takada et al, "Identification of Human TFIID Components and Direct Interaction Between a 250-kDa Polypeptide and the TATA Box-Binding Protein (TFIIDtau)", pages 11809-11813, see entire document.	1-4, 7-10, 13-21
Y	Proceedings of the National Academy of Sciences, USA, Volume 89, issued 1992, Timmers et al.,"Composition of Transcription Factor B-TFIID", pages 8140-8144, see entire document.	1, 2, 4, 7, 8, 10, 13-21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/01114

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C07H 17/00; C07K 7/04, 15/28; C12N 5/00, 15/00; C12P 21/06; C12Q 1/68

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 69.1, 172.3; 240.1, 320.1; 530/350, 388.1, 388.85; 536/23.1

Form PCT/ISA/210 (extra sheet)(July 1992)*

THIS PAGE IS BLANK